Regulation of virulence-related genes by RNA and RNA-interacting proteins in bacteria

DISSERTATION

zur Erlangung des akademischen Grades

Doctor of Philosophy
(Ph.D.)

eingereicht an der
Lebenswissenschaftlichen Fakultät der Humboldt-Universität zu Berlin

von
Andrés Escalera Maurer, MRes in Biochemical Research

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

Prof. Dr. Bernhard Grimm

Gutachter/innen
1. Prof. Dr. Kürsad Turgay
2. Prof. Dr. Emmanuelle Charpentier
3. Prof. Dr. Markus Landthaler

Tag der mündlichen Prüfung: 27 th May 2019

to Signe

to my family

Table of Contents

Acknowledgements

I would like to thank Manue for allowing me to pursue my interests and crazy ideas surrounded by very special, interesting and nice people. Yan Yan for being a grate example of commitment, passion and dedication. Shi Pey for redefining what hard-work and endurance means. Anaïs for being my unofficial, then official and then unofficial supervisor again, always being interested in my projects and correcting this thesis until the end (even when I asked her to stop). Also for the juggling, always spotting my typos and the writing together (despite the countless file versions). Anne-Laure for showing me that science can be a serious business, for the passionate (and not always focused) discussions and for saving me from some embarrassments in this thesis. Lina for always being there for me, willing to help with anything from making a Halloween costume to organizing my birthday parties. Eric for the long talks about science or life and for his concern in my well-being and survival. Inesita for making sure I didn't burn down the lab at the beginning (and at the end) and for her interest and effort in becoming more latina. Majda, who was the only person that was not annoyed by me, for being on my side (most of the times) and for being my Cas9 consultant. Franky for being a good friend despite being annoyed by me and for writing the abstract in German (thouh I can't judge the quality). Laurita for standing my repetitiveness, bad jokes and constant efforts to make her pissed and being a great companion in the office, lunch, dinner, beers, etc. Also for making me feel important and wise by asking me scientific questions of topics she knows more about than me. Victoria for being an amazing intern and bringing the SagA project back to life, working like crazy and spreading her motivation and joy. Also for making 'La Cantina' our second home. Vanessa for her eternal suggestions of new restaurants to try (within one block to her house) and sharing her eternal project with me in a way that made me feel part of it. Katja for making fun of my accent in 'German' (than means I know some words).

I also want to thank my TAC members, Kürşad and Markus, for their being tough when I needed it but always beeing interested, constructive and propositive.

I would need another thesis to thank all the people that have helped me and accompanied me during my PhD but I would like to thank all the members of the multiple Charpentier groups that I don't mention explicitly.

I would like to thank Signe for always supporting and believing in me and for not

basing her love in the success of my experiments. Also for being involved, up to date and informed about the lab gossip and keeping track of some of my deadlines (that of course I was aware of). To Marianne and Peter who helped with the several movings, if it was not for them I would not have running water in my home. Also for always making me feel welcome in my regenerating visits to Denmark.

My family for their unconditional support and confidence in me. My mum for being always interested in the developments of my PhD, visiting every time she could, integrating in the lab and even attending a Christmas outing. My father for his constant advice, specifically in handling conflicts and lab politics. My brother Manolo for allowing me to see problems from a different perspective or inviting me to escape from them in some other country when I needed it the most. My sister Gena for her interest and involvement in my career planning and encouraging me to continue. My brother Fernando for always being protective, concerned and suppotive of my decisions, despite having different wishes for my future.

Table of Figures

Abbreviations

AHL	acyl homoserine lactones
asRNAs	antisense RNAs
BLP	bacterial lipoprotein
Cas	CRISPR-associated proteins
CDS	coding sequence
CRISPR	clustered, regularly interspaced palindromic repeats
crRNA	CRISPR-RNA
EDTA	ethylenediaminetetraacetic acid
ES	early stationary
FMN	flavin mononucleotide
HPLC	high-performance liquid chromatography
ML	mid-logarithmic
NB	Northern blot
ncRNAs	non-coding RNAs
nt	nucleotides
O/N	overnight
OD	optical density
ORF	open reading frame

PAM	protospacer adjacent motif
PBS	phosphate buffer saline
pre-crRNA	pre-CRISPR RNA
PTS	Phosphoenolpyruvate phosphotransferase system
qRT-PCR	quantitative reverse transcription PCR
QS	quorum sensing
RBS	ribosomal binding site
RNase	ribonuclease
RNAseq	RNAseq
SAH	S-adenosylhomocysteine
SAM	S-adenosylmethionine
scaRNA	small-CRISPR-associated RNA
SLO	streptolysin O
SLS	streptolysin S
SP-STK	Ser/Thr kinase
SP-STP	Ser/Thr phosphatase
sRNA	small RNA
TCS	two-component system
THY	Todd Hewitt broth
TL	translational
TPP	thiamine pyrophosphate
tracrRNA	*trans*-activating CRISPR RNA
TSA	tryptic soy agar
TX	transcriptional
WT	wild type
YE	yeast extract

Abstract

Bacterial pathogens are constantly regulating the expression of their genes in response to changing environmental conditions and signals from the host. Timely and adequate levels of gene expression are essential for obtaining nutrients and evading the host immune system. The aim of this thesis was to study regulatory mechanisms of virulence-related genes in the bacterial pathogens *Francicella novicida* and *Streptococcus pyogenes.*

The focus of chapter one is on the regulation of the important virulence factor streptolysin S (SLS), which is responsible for the hemolytic phenotype of the human pathogen *S. pyogenes.* First, we investigated the role of the ribonuclease (RNase) Y in the transcriptional and post-transcriptional regulation of *sagA,* which codes for the precursor of SLS. We found that RNase Y promotes the production of a small RNA (sRNA) from the *sagA* transcript. However, no role of RNase Y in the regulation of the *sagA* transcript at the post-transcriptional level was observed. Yet, RNase Y promotes *sagA* transcription indirectly, affecting the hemolytic activity in a growth phase-dependent manner. Next, we studied the function of *sagA* 5′ untranslated region (UTR) as a putative *cis*-acting regulatory RNA. We show that the *sagA* 5′ UTR contains a secondary structure that may affect the accessibility to the ribosomal binding site (RBS) and that this structure is possibly modulated by direct binding to a ligand. Moreover, our results indicate that removing fragments of the 5′ UTR has a negative effect on *sagA* expression, possibly by stabilizing the RBS-blocking structure. While investigating the identity of the putative ligand that affects the *sagA* 5′ UTR structure, we developed a method for testing the activity of riboswitches. Using this method, we validated three predicted riboswitches in *S. pyogenes.*

In chapter two, we characterized the mechanism by which *F. novicida* CRISPR-Cas9 (FnoCas9) represses the expression of bacterial lipoproteins (BLPs), allowing evasion of the host immune system. We show that FnoCas9 is a dual-function protein that, in addition to its canonical DNA nuclease activity, evolved the ability to regulate transcription. In this newly-described mechanism, the non-canonical RNA duplex tracrRNA:scaRNA guides FnoCas9 to the DNA target located downstream of the promoter of the BLP-coding genes (FTN_1103 and FTN_1101), causing transcriptional interference. The endogenous targets contain a protospacer-adjacent motif (PAM) and a sequence that is complementary to scaRNA, promoting FnoCas9 binding. While the

mechanism is reminiscent of DNA targeting in the canonical immunity function of CRISPR-Cas9, with scaRNA fulfilling a similar function than crRNA, reduced complementarity between scaRNA and the DNA promotes binding but does not allow cleavage. This system can also be engineered to repress other genes, expanding the toolbox of CRISPR applications.

Zusammenfassung

Pathogene Bakterien passen ihre Genexpression konstant an sich verändernde Umweltbedingungen und Einflüsse des Wirtes an. Zeitlich abgestimmte und adäquate Genexpressionslevel sind essentiell für die Nährstoffaufnahme und um einer Immunantwort des Wirtes zu entgehen. Ziel dieser Arbeit war es, die regulatorischen Mechanismen von Virulenz-assoziierten Genen in den Pathogenen *Francisella novicida* und *Streptococcus pyogenes* zu untersuchen.

Kapitel eins befasst sich mit der Regulation des wichtigen Virulenzfaktors Streptolysin S (SLS), welcher für den hämolytischen Phänotyp des humanpathogenen Bakteriums *S. pyogenes* verantwortlich ist. Zunächst untersuchten wir die Funktion der Ribonuklease (RNase) Y während der transkriptionellen und posttranstrikptionellen Regulation des Gens *sagA*, welches für die Vorstufe von SLS kodiert. RNase Y begünstigte die Produktion einer kleinen RNA (*small RNA* – sRNA) vom *sagA* Transkript. Jedoch konnten wir keine Beteiligung der RNase an der posttranskriptionellen Regulierung des *sagA* Transkripts beobachten. Dennoch förderte RNase Y die Transkription von *sagA* indirekt, und damit, abhängig von der Wachstumsphase, die hämolytische Aktivität. Weiterhin untersuchten wir die Funktion der 5′-untranslatierten Region (UTR) des *sagA* Transkripts als ein putatives *cis*-wirkendes Element. Wir konnten zeigen, dass diese 5′ UTR eine Sekundärstruktur besitzt, die die Zugänglichkeit der ribosomalen Bindungsstelle (RBS) beeinflussen könnte, wobei die Struktur wahrscheinlich durch die Bindung eines Liganden moduliert wird. Außerdem deuten unsere Experimente darauf hin, dass die Deletion einzelner Abschnitte der 5′ UTR einen negativen Effekt auf die Expression von *sagA* hat, möglicherweise durch die Stabilisierung der RBS-blockierenden Struktur. Um den putativen Liganden zu identifizieren, der die Struktur der 5′ UTR von *sagA* beeinflusst, haben wir eine Methode entwickelt um die Aktivität von Riboswitches zu analysieren. Mit dieser Methode konnten wir drei putative Riboswitches in *S. pyogenes* validieren.

In Kapitel zwei charakterisierten wir den Mechanismus mit dem CRISPR-Cas9 aus *F. novicida* (FnoCas9) die Expression bakterieller Lipoproteine (BLPs) unterdrückt und damit dem Immunsystem des Wirtes entgeht. Wir konnten zeigen, dass FnoCas9 eine duale Funktion besitzt, die es dem Protein ermöglicht nicht nur DNA zu schneiden (kanonische Funktion), sondern auch Transkriptionsprozesse zu regulieren. Diese erstmals beschriebene Aktivität umfasst die Bindung von FnoCas9 an den nicht-kanonischen RNA-Duplex bestehend aus tracrRNA und scaRNA, wodurch der Protein-RNA Komplex an einen DNA Abschnitt stromabwärts des Promoters zweier BLP-kodierender Gene (FTN_1103 und FTN_1101) bindet und somit eine transkriptionelle Interferenz hervorruft. Diese endogene Bindungsstelle besitzt ein benachbartes Motiv (*protospacer-adjacent motif* – PAM) und eine scaRNA-komplementäre Sequenz, durch die der FnoCas9-RNA Komplex binden kann. Dieser Mechanismus erinnert an die kanonische DNA-bindende Immunfunktion von CRISPR-Cas9, wobei die scaRNA eine ähnliche Rolle wie die crRNA einnimmt. Jedoch begünstigt die verminderte Komplementarität zwischen scaRNA und der DNA zwar die Bindung, jedoch nicht die Spaltung der DNA. Dieses System kann auch dahingehend verändert werden, um die Expression anderer Gene zu reprimieren und erweitert damit das Repertoire an CRISPR-basierten Anwendungsmöglichkeiten.

Chapter One:
Regulation of streptolysin S expression by a small RNA and RNase Y in *Streptococcus pyogenes*

Regulatory RNAs in bacteria

Bacterial non-coding RNAs (ncRNAs) are involved in the regulation of central biological functions such as energy metabolism, quorum sensing, biofilm formation, stress response, adaptation to growth conditions and pathogenesis (Michaux et al., 2014). Traditional regulatory ncRNAs can be divided in four classes according to their mechanism of action: a) protein-activity modulation, b) antisense c) 5′-encoded regulatory elements (riboswitches and thermosensors) and, d) *trans*-encoded. Lately, CRISPR (clustered, regularly interspaced palindromic repeats) has emerged as an important new class of ncRNAs that are involved in defense against bacteriophage and plasmid invasion (see chapter two).

Trans-encoded ncRNAs base-pair with an mRNA target and activate or repress translation by diverse mechanisms. When the ncRNA base-pairs near or on the ribosome-binding site (RBS) of the target mRNA, it prevents the ribosome from binding to the mRNA and, therefore, inhibits translation. In other cases, the ncRNA base-pairs upstream of the RBS and promotes translation by inhibiting formation of secondary structures that, in the absence of the ncRNA, block access to the RBS. Additionally, *trans*-encoded ncRNAs can affect mRNA stability by promoting or inhibiting specific ribonuclease (RNase) activity. mRNA-ncRNA base-pairing can, for example, generate double-stranded (ds)RNA stretches that protect the RNA from single-stranded specific RNases. In other cases, ncRNA binding exposes single-stranded regions that, in the absence of the ncRNA would be double-stranded, allowing single strand(ss)-specific RNases to cleave. Of course, the opposite is also possible, i.e., ssRNA regions of the target mRNA that become ds when bound to the ncRNA, promoting ds-specific-RNases cleavage. Usually these kinds of ncRNAs interact via imperfect complementary sequences with their targets, allowing one ncRNA to have multiple targets (Storz et al., 2011).

Antisense RNAs (asRNAs) are *cis*-encoded ncRNAs transcribed from the

opposite strand of their targets and, consequently, fully complementary to them (Georg and Hess, 2018). asRNAs regulate gene expression by affecting transcription, mRNA stability or translation. mRNA stability and translation by *cis*-encoded sRNAs are regulated by similar mechanisms as those observed for *trans*-encoded ncRNAs. For asRNA-mediated transcriptional regulation two distinct mechanisms have been proposed: interference and attenuation. Interference means that simultaneous transcription from the sense and antisense strands cause the RNA polymerases to collide, interrupting the process. On the other hand, attenuation occurs when the asRNA causes the formation of a transcriptional terminator in the target mRNA (Sesto et al., 2013). In addition, ncRNAs can regulate protein activity either by acting as co-factors essential for protein activity or by antagonizing or sequestering proteins (Waters and Storz, 2009).

Finally, thermosensors and riboswitches regulate transcription or translation by changing the target RNA structure in response to changes in temperature or presence of a specific molecule, respectively (Ignatov and Johansson, 2017). Riboswitches will be described in more detail in the following section.

Riboswitches

Riboswitches are RNA structures that specifically bind small molecules and modify gene expression. Typically, riboswitches are found in the 5′ UTR of mRNAs but can also be present in ncRNAs such as as and protein-sequestering RNAs (DebRoy et al., 2014; Mellin et al., 2013, 2014). Furthermore, some riboswitch-containing transcripts also act as *trans*-encoded ncRNAs (Loh et al., 2009). Known riboswitch ligands include ions, cofactors (e.g. vitamins) and modified nucleotides (nt) (such as second messengers). Riboswitches are widely distributed in bacteria and can also be found in some fungi, algae and plants (Barrick and Breaker, 2007; Breaker, 2012). To date, approximately 40 structurally distinct riboswitch classes have been discovered (Lotz and Suess, 2018). Even though each riboswitch class senses a specific ligand, some ligands can be sensed by more than one riboswitch class (Lotz and Suess, 2018).
Riboswitches consist of two elements: a ligand-sensing domain, known as an aptamer, and a regulatory domain, called expression platform. In response to changes in ligand concentration the expression platform undergoes a conformational change that regulates expression of the downstream transcript. Regulatory mechanisms of

riboswitches include modulation of transcription, translation, transcript stability and processing (Barrick and Breaker, 2007). The expression platform of transcriptional riboswitches can adopt two mutually exclusive conformations: an intrinsic transcriptional terminator that prevents transcription elongation, or an anti-terminator that allows transcription to continue. Similarly, translational riboswitches take two alternative structures that permit or block access to the RBS. Other, less-studied riboswitch mechanisms involve Rho-dependent transcriptional termination and modulation of ribonucleolytic processes, either by self-cleaving ribozymes or ribonucleases (Hollands et al., 2012; Lee et al., 2010; Winkler et al., 2004).

In some cases, multiple regulatory mechanisms can be integrated in one expression platform to give rise to more complex systems, for example by simultaneously regulating translation initiation and cleavage by an RNase (Caron et al., 2012). Ligand binding usually inhibits expression of the adjacent gene although upregulation has also been reported (Mandal and Breaker, 2004; Sudarsan et al., 2008).

Discovery of riboswitches and ligand identification

In order to fulfil their function, aptamers need to specifically recognize their ligand at physiological concentrations (in the pM to mM range, depending on the riboswitch) and discriminate between very similar molecules. This imposes constraints at the level of structure and sequence, which makes the aptamer the most conserved part of riboswitches (Breaker, 2011). This aptamer conservation has been successfully exploited to predict riboswitches (Ames and Breaker, 2010). However, bioinformatics approaches have thus far focused on structures that are widely distributed across species (Barrick et al., 2004; Weinberg et al., 2007). It is therefore likely that riboswitches with narrower distributions have been overlooked. In many cases, the identity of the ligand has been inferred based on the genetic context (Barrick et al., 2004). Yet, finding the ligand in cases where the function of the adjacent gene is unknown can be challenging (Meyer et al., 2011). Moreover, small sequence variations of even a single substitution can alter the specificity of the riboswitch, making it difficult to predict the identity of the ligand even for closely related riboswitches (Weinberg et al., 2017). Therefore, individual riboswitch variants still need to be experimentally

validated.

Traditionally, a technique called in-line probing has been used to evaluate the binding of ligands to their corresponding riboswitch (Regulski and Breaker, 2008). In-line probing takes advantage of the property of RNA to spontaneously self-cleave in a structure-dependent manner, with single-stranded regions being more prone to degradation (Regulski and Breaker, 2008). Therefore, denaturing-gel electrophoresis analysis of 5′-end labelled RNA after incubation with putative ligands results in a band pattern that provides structural information. Further analysis can even help estimating differences in affinity between closely related molecules (Regulski and Breaker, 2008). In-line probing has also been used to identify the ligand of a predicted riboswitch from a complex mix of metabolites (Nelson et al., 2013). However, this approach is laborious and, since it renders no functional information, can lead to false positive hits.

Currently, no high-throughput method for identifying ligands of predicted riboswitches exists. In addition to aiding the discovery of endogenous ligands of riboswitches, such a method could be used to identify non-natural ligands of known riboswitches. This knowledge could be harnessed for the development of new antibiotics (Aghdam et al., 2016).

Streptococcus pyogenes

Streptococcus pyogenes is a Gram-positive bacterium that is only known to infect humans. It forms chains of *cocci* and causes the complete lysis of red-blood cells (beta-hemolysis). Colonization by *S. pyogenes* can have a wide variety of outcomes, from asymptomatic carriage and mild local colonization in the skin or throat, to deep-tissue and systemic invasions (bacteremia). Pharyngitis (sore throat) is the most frequent disease caused by *S. pyogenes.* In addition, *S. pyogenes* is the predominant non-viral cause of pharyngitis (Wessels, 2016). Throat infection, and other streptococcal diseases, can be accompanied by scarlet fever, a skin rash that is likely caused by exposure to streptococcal toxins (Wessels, 2016).

S. pyogenes can also infect different skin layers, causing impetigo or erysipelas when the infection is at the superficial keratin layer and epidermis, respectively (Stevens and Bryant, 2016). Infection of the deeper tissue can lead to more severe diseases such as necrotizing fasciitis, which can have a mortality rate of up to 80 % (Stevens and Bryant, 2016). Superantigens and other virulence factors produced by *S. pyogenes* may cause an excessive immune response resulting in streptococcal

toxic shock syndrome and organ failure (Stevens and Bryant, 2016).

Finally, cross-reactivity with the antigens that are present on the surface of *S. pyogenes* can also lead to post-streptococcal autoimmune sequelae such as acute rheumatic fever leading to rheumatic heart disease or post-streptococcal glomerulonephritis (Cunningham, 2016).

Hemolysins in *Streptococcus pyogenes*

S. pyogenes secretes multiple virulence factors that help the bacteria obtaining nutrients and in the defense against the immune system of the host (Hynes and Sloan, 2016). Among the most studied virulence factors are the cytolysins, streptolysin S (SLS) and streptolysin O (SLO).

Streptolysin O

SLO is an oxygen-labile pore-forming cytotoxin that is translated as a 69 kDa protein which is activated by a proteolytic cleavage and exported to the extracellular milieu (Hynes and Sloan, 2016). The mature SLS is then inserted in the membrane of host cells in a cholesterol-dependent manner and oligomerizes forming a pore (Hynes and Sloan, 2016). In addition to cholesterol, a galactose-containing receptor is involved in SLO-mediated pore formation in some conditions (Mozola and Caparon, 2015; Shewell et al., 2014). In macrophages, pore formation leads to caspase-dependent apoptosis (Timmer et al., 2009). Consistently, SLO negative mutants are less resistant to killing by macrophages when compared to the isogenic SLO positive strains (Bastiat-Sempe et al., 2014), and are attenuated in virulence (Fontaine et al., 2003; Limbago et al., 2000). Following phagocytosis, SLO prevents acidification, allowing the bacteria to survive (Bastiat-Sempe et al., 2014). SLO also activates neutrophils (Nilsson et al., 2006), promotes inflammation and boosts the immune response (Harder et al., 2009). SLO is immunogenic and has been proposed as a potential candidate for vaccine development (Chiarot et al., 2013).

Streptolysin S

SLS is an oxygen-stable thiazole/oxazole-modified microcin toxin produced by *S. pyogenes* and other streptococcal species (Molloy et al., 2011). The genes that are necessary for the production and secretion of SLS are encoded in the nine-gene

operon *sag* (Nizet et al., 2000). The first gene of the operon, *sagA,* encodes the SLS precursor peptide that is modified and exported by the remaining Sag proteins. The genes *sagBCD* code for a trimeric oxazole/thiazole synthase complex (SagBCD) that modifies SagA conferring it cytolytic activity (Lee et al., 2008). The SagE protease removes the leader of the modified SagA, giving rise to the active SLS toxin (Maxson et al., 2015). Once modified, the active SLS is exported via an ABC transporter formed by the SagGHI proteins (Datta et al., 2005). The remaining Sag protein (SagF) has an unknown function but it is also essential for hemolytic activity (Nizet et al., 2000).

Role in Virulence

The *sag* operon is conserved across almost all studied strains (Nizet et al., 2000; Yoshino et al., 2010) suggesting it is important for the survival of *S. pyogenes*. Indeed, mutant strains that are unable to produce SLS are attenuated in virulence and cause less tissue damage in most murine models of infection, compared to their corresponding isogenic wild type (WT) strain (Betschel et al., 1998; Datta et al., 2005; Engleberg et al., 2004). The contribution of SLS to virulence varies depending on the model and the studied strain, with some models showing little contribution to survival or pathogenicity (Fontaine et al., 2003; Kinkel and McIver, 2008). In some strains, the relative contribution of SLS to pathogenesis varies in different strains depending on the expression of other factors such as the capsule (Sierig et al., 2003). It was shown that *sagA* deletion mutant is attenuated in a murine invasive model only when the strain is also unable to produce capsule (Sierig et al., 2003).

Despite the limitations of current infection models for *S. pyogenes* (Watson et al., 2016), it is now widely accepted that SLS is an important virulence factor for *S. pyogenes* (Hynes and Sloan, 2016). However, the specific functions of SLS during infection are less clear. The proposed roles of SLS include defense against the immune systems of the host, dissemination across tissues, and ensuring nutrient availability (Molloy et al., 2011).

The implication of SLS in defense against the immune system of the host is supported by evidence showing that SLS mediates neutrophil and macrophage killing. Indeed, it was shown that *S. pyogenes* cytotoxicity on macrophages is mostly mediated by SLS and SLO (Goldmann et al., 2009). It has also been observed that *S. pyogenes* kills neutrophils in an SLS-dependent manner (Miyoshi-Akiyama et al., 2005). In addition to its observed cytotoxicity, SLS is able to inhibit neutrophil recruitment to the

site of infection (Feng et al., 2017; Lin et al., 2009). Consequently, an SLS-deficient mutant was attenuated in virulence and was associated with an increased accumulation of neutrophils compared to the isogenic WT strain (Feng et al., 2017; Lin et al., 2009). In agreement with these results, a recent study has found that SLS activates pain-sensing neurons, which in turn block neutrophils recruitment (Pinho-Ribeiro et al., 2018).

Apart from its role in defense, SLS has been suggested to facilitate the dissemination of *S. pyogenes* across different tissues. Accordingly, the ability of *S. pyogenes* to translocate across epithelial cells *in vitro* was reduced in a SLS negative mutant compared to the WT (Sumitomo et al., 2011). Interestingly, SLS acts indirectly via the host protease calpain to mediate proteolytic cleavage of intercellular junctions (Sumitomo et al., 2011). A recent study has also linked SLS and SLO with biofilm production in cell cultures and microcolony formation in a mouse model of necrotising fasciitis (Vajjala et al., 2018). This study shows that this is dependent on the ability of the streptolysins to cause endoplasmic reticulum stress and proposes that this promotes biofilm formation, dissemination and proliferation indirectly through the release of unknown signals (Vajjala et al., 2018).

Mode of action

Despite the fact that SLS has been known to lyse cells since 1938 (Molloy et al., 2011), the precise mechanism remains largely unknown. The most detailed biochemical study so far shows that the interaction between SLS and the ion transporter Band 3 mediates lysis of red blood cells by facilitating influx of Cl^- (Higashi et al., 2016). Furthermore, inhibition of Band 3 activity reduces skin lesion size to similar levels than deleting *sagA* in a murine model of skin infection (Higashi et al., 2016). However, since the expression of the Band 3 protein is restricted to erythrocyte, the mechanism that mediates SLS-dependent lysis in other cell types is currently unknown.

Regulation

As mentioned above, SLS is an important virulence factor in *S. pyogenes*. As such, the conditions in which this toxin is produced have been broadly studied. It is important to note that due to inter-strain variability it is impossible to make general conclusions about the role some of these factors have on SLS regulation. Furthermore, even if the

regulators themselves are conserved, their regulon might vary in different strains and/or conditions. However, because the cues and pathways that affect SLS production are likely related to its function, some general conclusion can be drawn out of this information.

The complete signal transduction pathway linking the input signal to changes in SLS production has not been traced in most cases. Yet, various conditions and cues that affect *sagA* expression (and some of its regulators) have been discovered. These include nutrient availability (e.g. glucose and nitrogen), growth in blood, saliva or conditioned media and presence of small molecules (such as homoserine lactones, asparagine and SLS autoregulation) (Baruch et al., 2014; Graham et al., 2005; Salim et al., 2007; Saroj et al., 2016, 2017; Shelburne et al., 2010; Steiner and Malke, 2001; Sundar et al., 2018; Valdes et al., 2018). Some of the factors that modulate SLS production include stand-alone regulators (e.g. Mga, CcpA), two-component systems (e.g CovR/S, Ihk/irr, SptR/S), RNases (i.e. RNases Y, J1, J2, PNPase) and a sRNA (fasX) (Vega et al., 2016).

In spite of the body of knowledge that has accumulated regarding conditions that affect SLS production, there is little information about the mechanisms governing the transcriptional regulation of *sagA* and even less about the factors affecting SLS production at the post-transcriptional level. The cases where the specific signal that is sensed is known or the regulatory mechanism has been elucidated, are explained in more detail below.

Regulation by small molecules and quorum sensing

Bacteria rely on the production and detection of small molecules in order to sense the presence and abundance of other bacteria in the surrounding environment, a system called quorum sensing (QS).

Sil is a QS system composed of the SilAB two-component system (TCS), the SilDE ABC transporter and the SilCR signalling peptide (Hidalgo-Grass et al., 2002). Between 12% and 25% of *S. pyogenes* isolates encode Sil, with some bacteria having incomplete or non-functional systems (Jimenez and Federle, 2014). It has been shown that the pheromone SilCR upregulates *sagA* expression (Salim et al., 2008). Interestingly, this effect was observed even in absence of SilAB suggesting the presence of other mechanisms to sense SilCR from other strains or species, even in the absence of the complete Sil QS system (Salim et al., 2008).

Other QS signaling molecules have been recently implicated in *sagA* regulation. Acyl homoserine lactones (AHLs), typically involved in bacterial QS systems, were shown to enter *S. pyogenes* cells through the ferrichrome transporter FtsABCD and repress *sagA* expression (Saroj et al., 2017). Regulation of *sagA* by AHLs is dependent on the QS transcriptional regulator LuxR that was shown to bind *sagA* promoter (Saroj et al., 2017). However, the exact mechanism mediating this regulation is unclear, as LuxR also seems to bind *sagA* promoter in the absence of AHLs, at least *in vitro*. In addition, the inhibitory effect of AHLs was not observed in all the strains studied, suggesting that it might be strain-specific (Saroj et al., 2017). Interestingly, the same study detected an increase in the intracellular iron concentration after addition of AHLs and proposed that inhibition of *sagA* expression is mediated by iron (Saroj et al., 2017). Though these hypotheses need further investigation, they are in line with the proposed role of SLS in iron acquisition (Molloy et al., 2011).

In contrast, a previous study found that *sagA* expression was upregulated in high (1000 μM) compared to low (1 μM) iron concentrations (Salim et al., 2007), which is in agreement with the upregulation of *sagA* in blood (Graham et al., 2005). The authors proposed that high iron concentrations mimic the environment inside the host phagosome and SLS production allows *S. pyogenes* to escape (Salim et al., 2007). It is therefore possible that different iron concentrations, or iron signalling under different conditions, have opposing effect.

In addition to AHL, SLS itself has been shown to act as a QS signal via an unknown mechanism (Salim et al., 2007). Conditioned media from WT *S. pyogenes* but not from a *sagA* deletion mutant induced *sagA* expression (Salim et al., 2007). The same effect was observed upon addition of purified SLS to the medium (Salim et al., 2007). This is in contrast to a previous study showing that addition of conditioned media had no effect on *sagA* expression (Mangold et al., 2004). Therefore, whether SLS acts as a QS molecule or whether it is strain specific remains unclear.

The amino acid asparagine is the only other example where the concentration of a specific molecule is linked to *sagA* regulation. A study by Baruch and colleagues has found that depletion of asparagine induces expression of the *sag* operon partly through the TrxRS TCS (Baruch et al., 2014). In addition, SLS and SLO cause endoplasmic reticulum stress, leading to the production of asparagine. Since asparagine promotes *S. pyogenes* growth *in vitro*, the authors proposed that one of

the functions of SLS and SLO is to induce the release of asparagine in order to favour growth (Baruch et al., 2014).

Direct transcriptional regulation

Of all the known transcriptional regulators that affect *sagA* expression, only two in addition to LuxR (which seems to act in strain-specific manner) have been shown to bind the *sagA* promoter region: CcpA and CovR.

CcpA is the catabolite control protein that regulates carbohydrate utilization via the phosphoenolpyruvate phosphotransferase system (PTS), which monitors availability of different carbon sources (Deutscher et al., 2006). It was shown that CcpA represses *sagA* in response to carbon catabolite repression (DebRoy et al., 2016; Kietzman and Caparon, 2010; Kinkel and McIver, 2008; Shelburne et al., 2008). However, there is contradictory evidence as to whether this regulation is direct or indirect. While Kietzman and Caparon, 2010 saw no interaction between CcpA and a putative CcpA-binding site upstream of the *sagA* promoter, others have observed a direct interaction (Kinkel and McIver, 2008; Shelburne et al., 2008). Nonetheless, the regulation of *sagA* by CcpA is conserved across multiple serotypes (DebRoy et al., 2016) and might explain, at least in part, the repression of *sagA* expression in the presence of glucose (Sundar et al., 2018; Valdes et al., 2018).

CovR/S is the TCS that controls the expression of several virulence factors. CovS responds to Mg^{2+} and host antimicrobial peptides (Gordon, 2007; Gryllos et al., 2003, 2008) and phosphorylates the response regulator CovR. Phosphorylated CovR represses *sagA* expression by binding two sites located in the vicinity of the *sagA* promoter (Gao et al., 2005; Horstmann et al., 2014). Interestingly, CovR does not require CovS to regulate *sagA* (Dalton and Scott, 2004; Horstmann et al., 2014), suggesting there are other mechanisms for CovR phosphorylation. Indeed, CovR reversible phosphorylation by the Ser/Thr kinase (SP-STK) and phosphatase (SP-STP) has been shown to affect *sagA* expression (Agarwal et al., 2011).

Post-transcriptional regulation

In addition to the effect that transcriptional regulators have on *sagA* expression, there is some evidence, albeit scarcer, indicating that production of SLS might also be regulated at the post-transcriptional level. Four RNases have been shown to affect

sagA transcript abundance and/or stability (PNPase, RNase Y and RNases J1/J2) (Barnett et al., 2007; Bugrysheva and Scott, 2010; Kang et al., 2010).

In 2007, Barnett and colleagues proposed that two factors are responsible for the increase in *sagA* transcript abundance at early stationary (ES) growth phase as compared to mid-logarithmic (ML) (Barnett et al., 2007). The first is an increment in promoter activity and the second a stabilization of the *sagA* transcript (Barnett et al., 2007). They further discovered that the 3′ to 5′ exoribonuclease polynucleotide phosphorylase (PNPase) is involved in decay of the *sagA* transcript (Barnett et al., 2007). Indeed, while deletion of two other 3′ to 5′ exoRNases (RNase R and YhaM) had no effect in transcript stability, *sagA* mRNA was 8-fold more stable in a mutant lacking PNPase (Barnett et al., 2007). Though these results do not necessarily mean that PNPase is involved in regulating hemolysis, the difference in *sagA* stability between the two growth phases suggest that SLS production might be regulated at the transcriptional and post-transcriptional levels.

Other RNases involved in the *sagA* mRNA degradation are RNases J1 and J2, which are essential in *S. pyogenes* (Bugrysheva and Scott, 2010). RNase J1 is the only described 5′-to-3′ exoRNase in bacteria and might act as an endoribonuclease in some cases, though this latter activity is still under debate (Durand and Condon, 2018). RNase J2 is an orthologue of RNase J1 whose activity is less understood but seems to form a complex with RNase J1 (Durand and Condon, 2018). Using conditional mutants of RNases J1 and J2, Bugrysheva and Scott show that the decay of *sagA* transcript initiates earlier when the expression of the RNases is induced (Bugrysheva and Scott, 2010). These results indicate that RNases J1 and J2 might be involved in the turnover of *sagA* transcript (Bugrysheva and Scott, 2010).

RNase Y is a single-stranded specific endoRNase that is anchored to the inner side of the membrane (Durand and Condon, 2018). This enzyme is important for the virulence in various Gram-positive bacteria including *S. aureus*, *C. perfringens*, and *S. pyogenes* (Chen et al., 2013; Kaito et al., 2005; Kang et al., 2010; Khemici et al., 2015; Marincola et al., 2012; Nagata et al., 2008; Obana et al., 2017). In *B. subtilis,* RNase Y has an impact on the transcript abundance of most riboswitches (DeLoughery et al., 2018) and other *cis*-acting RNA structures (Laalami et al., 2013). It has been shown that RNase Y cleaves the SAM-binding riboswitch, preferably in the presence of the ligand (Shahbabian et al., 2009). This suggests that, at least in *B. subtilis,* RNase Y is a key player in riboswitch-mediated regulation.

The biochemical constraints that determine the specificity of RNase Y remain largely unknown. In *S. aureus* RNase Y was reported to cleave preferably after G in A/U-rich regions (Khemici et al., 2015) and the requirement for a secondary structure downstream of the cleavage site was proposed (Marincola and Wolz, 2017). A recent study from our laboratory showed that a presence of a G is required for RNase Y cleavage but failed to identify any structural requirement (Broglia et al., 2018).

In *S. pyogenes,* RNase Y was reported to regulate the expression of approximately 29% of the genome, including *sagA,* which was downregulated in a mutant strain unable to produce RNase Y compared to the WT (Kang et al., 2010). Although these studies open the possibility that these RNases are involved in the post-transcriptional regulation of *sagA,* the mechanism and the extent of these effects remain to be investigated.

Results

RNase Y is involved in the processing of *sagA* transcript

A previous study from our laboratory that used RNA sequencing (RNAseq) and Northern blot analyses to discover novel sRNAs in *S. pyogenes* showed that *sagA* mRNA contains a 144 nt-long 5′ UTR that gives rise to a sRNA (Rhun et al., 2016).

Because RNase Y was reported to regulate *sagA* expression (Kang et al., 2010), it is possible that RNase Y cleaves the *sagA* 5′ UTR thus regulating *sagA* expression. Indeed, a ~120 nt-long sRNA was detected in the WT and the Δ*rny* complemented (Δ*rny::rny*) strains but not in the Δ*rny* strain by Northern blot analyses (Figure 1. A-B). Interestingly, even though RNase Y is produced at similar levels throughout the growth phases (Lécrivain, unpublished), this sRNA was observed in ML but not in ES growth phases (Figure 1. B), suggesting that it is produced by a regulated process. Therefore, we first investigated the mechanism by which the *sagA* transcript was processed and the exact location of the processing site.

Several attempts to determine the exact position of the cleavage by primer extension were unsuccessful (data not shown). It is possible that the downstream fragment produced by RNase Y processing was too unstable to be detected by primer extension. Indeed, cleavage products were not detected with a probe that anneals to the *sagA* coding sequence (CDS) (Figure 1. C). To overcome these limitations, we

performed RNA circularization followed by reverse transcription-PCR of the sRNA and sanger sequencing. Yet, transcript ends corresponding to the end of the sRNA were not detected (data not shown). A possible explanation is that the ratio between processed and unprocessed transcripts is low, reducing the probability of detecting the processing site by this technique. In support of this, RNA stability assays showed that the *sagA* primary transcript was highly stable in the WT and in the Δ*rny* mutant, in both growth phases (Figure 1. D). Because processing of the primary transcript would appear as a reduction of band intensity, this suggested that the rate at which the primary transcript was processed is low.

As an alternative approach to determine the exact location of RNase Y cleavage, we generated deletions and point mutations in the area surrounding the RNase Y putative cleavage site. First, we estimated the approximate location of the 3′ end using the size of the sRNA in the Northern Blot. Then, we constructed a transcriptional reporter fusion to the firefly luciferase gene expressing *sagA* 5′ UTR under the P23 constitutive promoter (P23-5′ UTR) and introduced substitutions in the RNase Y putative cleavage site (Figure 1. A). Because it was recently reported that RNase Y requires a guanosine (G) adjacent to the cleavage site to be active (Broglia et al., 2018), we introduced G-to-A substitutions in all Gs in this area (Figure 1. A). In addition, we deleted up to 10 nucleotides surrounding the putative cleavage site (Figure 1. A). However, both the substitutions and the deletions failed to inhibit the production of the sRNA (Figure 1. E). Preliminary results suggested that the reporter fusion that contains the WT sequence gives rise to the sRNA in Δ*rny* as well as in the WT (data not shown). Though these results need to be confirmed, this raises the possibility that the sRNA is produced by other mechanisms and that RNase Y affects its production indirectly.

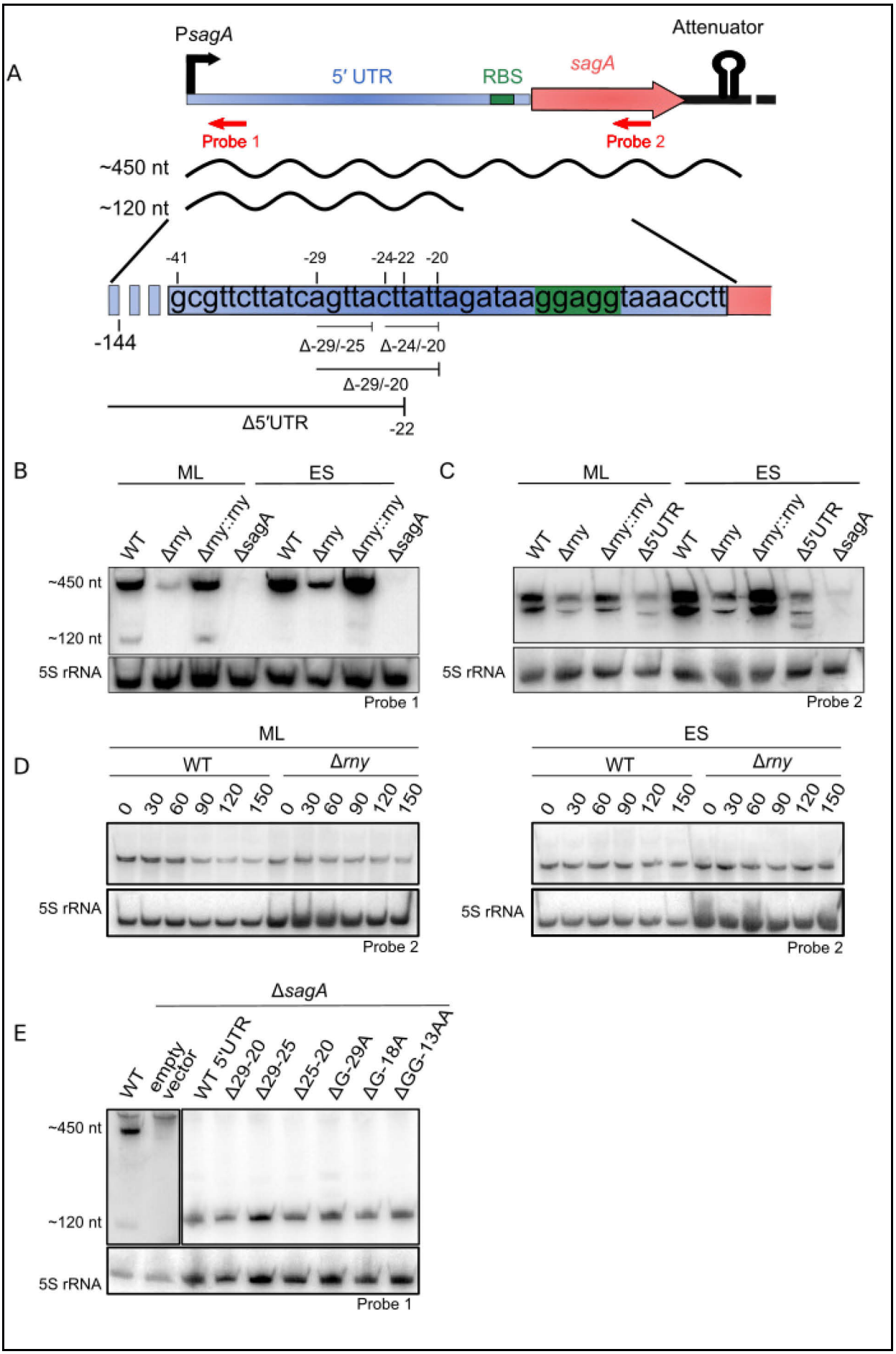
A
PsagA
Attenuator
5′ UTR
RBS
sagA
Probe 1
Probe 2
~450 nt
~120 nt
-41
-29
-24-22
-20
gcgttcttatcagttacttattagataaggaggtaaaacctt
-144
Δ-29/-25
Δ-24/-20
Δ-29/-20
Δ5′UTR
-22
B
ML
ES
WT
Δrny
Δrny::rny
ΔsagA
~450 nt
~120 nt
5S rRNA
Probe 1
C
ML
ES
WT
Δrny
Δrny::rny
Δ5′UTR
ΔsagA
5S rRNA
Probe 2
D
ML
ES
WT
Δrny
0
30
60
90
120
150
5S rRNA
Probe 2
E
ΔsagA
WT
empty vector
WT 5′UTR
Δ29-20
Δ29-25
Δ25-20
ΔG-29A
ΔG-18A
ΔGG-13AA
~450 nt
~120 nt
5S rRNA
Probe 1

Figure 1. RNase Y affects expression and processing of the *sagA* transcript. A) Schematic representation of the *sagA* locus (top). Annealing sites of Northern blot probes are indicated with red arrows. Undulated lines represent the transcripts detected by Northern Blot, approximate sizes of transcripts are indicated. The bottom panel shows a close-up of the region surrounding the RBS indicating the sequence and the position of deletions/substitutions introduced in the *sagA* 5′ UTR. Numbers indicate coordinates relative to the *sagA* start codon. B and C) Northern blot analysis of *sagA* transcript during mid-logarithmic (ML) and early stationary (ES) growth phases in wild-type (WT), RNase Y deletion (*Δrny*), complementation (*Δrny::rny*) strain and 5′ UTR *sagA* deletion (Δ5′UTR) strain. *sagA* deletion strain (*ΔsagA*) was used as a control to confirm probe specificity. D) Stability assay of *sagA* transcript in WT and *Δrny* grown until ML (left panel) or ES (right panel) growth phases. Transcription was stopped using rifampicin and samples were taken as indicated. Numbers indicate time points in minutes. Approximately 3-fold more RNA was used for *Δrny* than for WT to compensate for lower initial transcript abundance. E) Northern Blot analysis of Δ*sagA* transformed with the empty vector (pEC2174), the vector containing the WT *sagA* 5′ UTR or the 5′ UTR with the mutations described in B, the WT strain was included as a control. The 5S rRNA was used as a loading control for all Northern blots. In E, the size of the 5S RNA band is approximately the same as the band for the sRNA, which interferes with the loading control making it impossible to make any conclusion about the abundance of the sRNA. However, the presence or absence of the sRNA can be evaluated. The probes used in each experiment is indicated at the bottom. Probe 1 (OLEC3273) anneals at positions −122 to −100 of *sagA* 5′ UTR (see also A). Probe 2 (OLEC7883) is complementary to the last 26 nt of *sagA* CDS (see A).

RNase Y regulates *sagA* mRNA expression

As mentioned above, RNase Y regulates *sagA* transcript abundance (Kang et al., 2010). In agreement, *sagA* transcript levels were 4-fold lower in Δ*rny* compared to WT in ML growth phase (Le Rhun, RNAseq differential expression analysis unpublished). Furthermore, Northern Blot analyses showed that the abundance of *sagA* primary transcript was lower in Δ*rny* than in the WT or the complemented (Δ*rny::rny)* strain, regardless of the growth phase (Figure 1. B-C). Production of the sRNA should result in lower primary transcript levels, therefore the RNase Y-dependent upregulation of the primary transcript must result from a different process. This indicates that RNase Y has two opposing effects on the abundance of primary *sagA* transcript (one at the transcriptional and one at the post-transcriptional levels).

In order to investigate the contribution of the post-transcriptional effect, we analyzed the expression of the constitutive P23-5′UTR fusion in the WT and Δ*rny* (Figure 2. A). However, no difference in expression was observed between the two strains in either ML and ES growth phases (Figure 2. A). This is in agreement with the

fact that the abundance of the processed transcript is low in comparison to the primary transcript (Figure 1. B), suggesting that it was produced at a slow rate.

It is possible that under conditions where the processing rate increases, the post-transcriptional effect is stronger. To investigate the effect that removing *sagA* 5′ UTR had on transcript abundance, we used the Δ5′ UTR strain, which contained a deletion of the first 122 nt of *sagA* 5′ UTR in the chromosome of *S. pyogenes* (leaving 22 nt upstream of the start codon, Figure 1. A). Northern blot analyses showed a lower *sagA* mRNA abundance in Δ5′ UTR compared to the WT strain, in both growth phases (Figure 1. A). This suggested that a processing event that removes the *sagA* 5′ UTR from the transcript would have a negative impact on *sagA* transcript abundance.

To investigate the role of RNase Y in the transcriptional regulation of *sagA* independently of any post-transcriptional effect, we constructed a fusion of *sagA* 5′ UTR to the firefly luciferase reporter gene containing the P*sagA* promoter without the *sagA* 5′ UTR and tested its expression in *S. pyogenes* (Figure 2. B, PsagA-fflux). Unexpectedly, Δ*rny* and WT strains showed similar expression levels for these constructs in both ML and ES growth phases (Figure 2. B). It is possible that the reporter system does not recapitulate the natural regulation due to plasmid copy number or other artefacts.

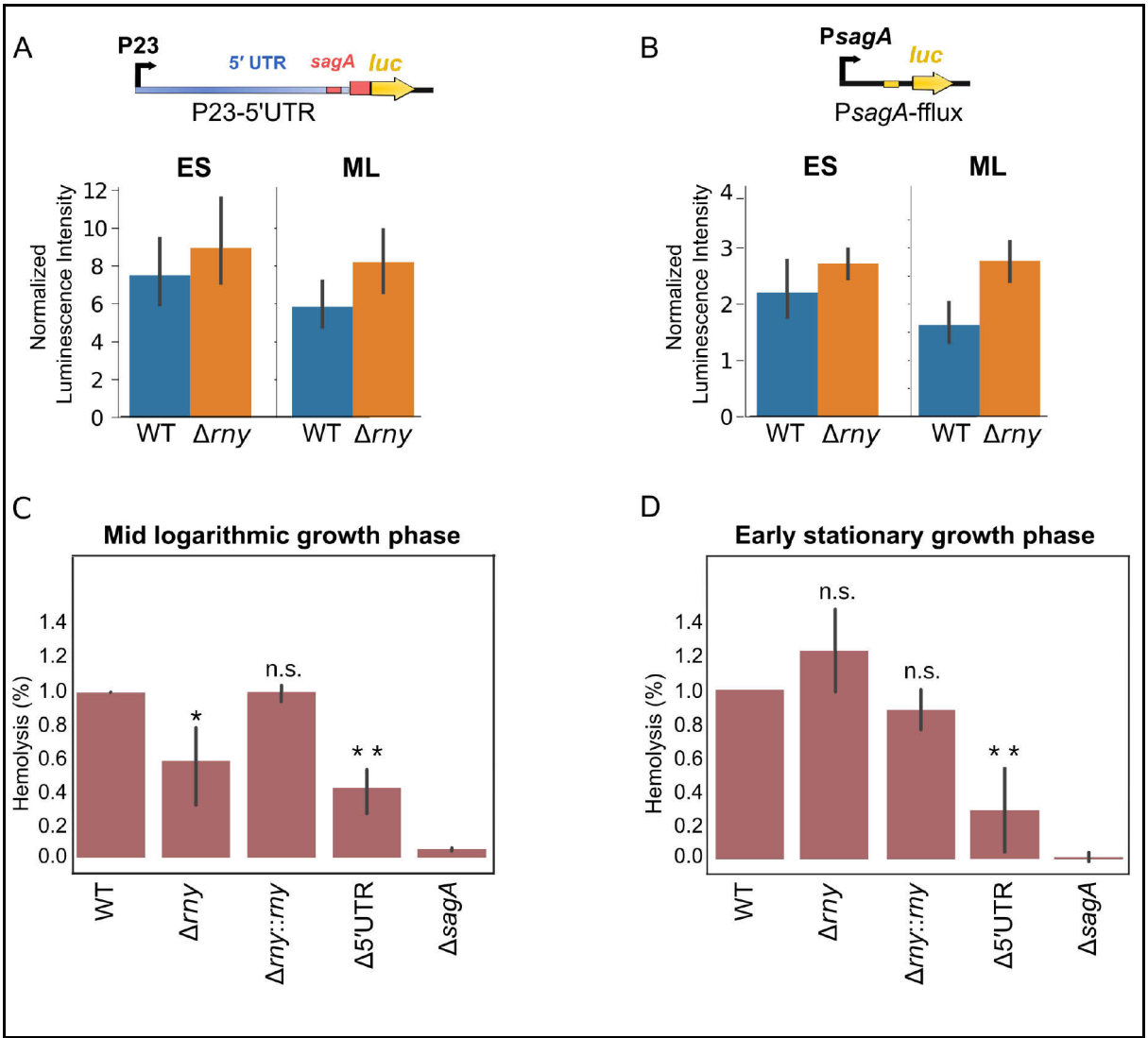

Figure 2. Hemolytic activity is reduced in *Δrny* compared to the WT strain. Activity of reporter fusions containing either the P*sagA* promoter A) or Δ5′ UTR *sagA* under the P23 constitutive promoter B) in WT (blue bars) and Δ*rny* (orange bars) cultured until ML or ES growth phases. Luminescence intensity was normalized against the control plasmid (pEC2174, containing the P23 constitutive promoter). Bars show averages for at least three independent biological replicates, error bars represent standard deviations. Streptolysin S (SLS)−dependent hemolytic activity of WT, *Δrny* and *Δrny::rny* and Δ5′ UTR *sagA* in ML (C) and ES (D) growth phases. *ΔsagA* was used as a control. The average of three independent biological replicates as percentage of the activity of the WT strain is shown. Error bars indicate standard deviation. Independent t−test *p* values are indicated for relevant comparisons: n.s = $p > 0.05$, * = $p \leq 0.05$, ** = $p \leq 0.01$.

RNase Y deletion affects hemolysis in ML but not in ES growth phase

It is clear that the deletion of RNase Y has a negative effect on *sagA* transcript abundance. In order to test whether RNase Y regulation of *sagA* transcript has an impact in SagA production, we cloned *sagA* fused to a Flag tag in a plasmid. However, after several attempts, no signal was detected corresponding to an expressed SagA-Flag in *S. pyogenes* WT and Δ*sagA* strains containing this recombinant plasmid by Western Blot (data not shown).

As an indirect approach to detect SagA production, we measured the hemolytic activity of Δ*rny* and WT. As expected, the hemolytic activity of Δ*rny* cultures in ML growth phase was significantly lower when compared to WT or Δ*rny::rny* strains (Figure 2. C). Surprisingly, no difference was observed when the hemolysis assay was carried out using cultures in ES growth phase (Figure 2. D). Because transcript expression is lower in the absence of RNase Y in both growth phases, these results indicate that SLS production is uncoupled from transcript abundance, suggesting that there are additional mechanisms regulating SLS production. In contrast, hemolytic levels were lower in the Δ5′ UTR *sagA* compared to the WT strain, in both growth phases (Figure 2. C-D), indicating that lower transcript levels are not always compensated by other processes.

Further experiments are needed in order to understand the contribution of RNase Y to the transcriptional and post-transcriptional regulation of *sagA* expression. Nevertheless, RNase Y seems to affect the *sagA* transcript by two potentially independent mechanisms i) inducing the production of a sRNA from the 5′ UTR of the *sagA* transcript, ii) upregulating *sagA* transcription through an unknown intermediate factor.

Truncations of *sagA* 5′ UTR affect *sagA* expression levels

As shown above, the *sagA* 5′ UTR is required for WT-levels of *sagA* expression and SLS production. In order to investigate the regions (and structures) that are important for SagA production, we generated reporter fusions containing various truncations on the 5′ end of the *sagA* transcript (Figure 3. A). The expression of the truncated fusions was evaluated using the mVenus fluorescent protein as reporter in *E. coli* (Figure 3. B-D).

A first set of translational (TL) fusions included, in addition to the 5′ UTR fragment, the first 54 nt of the *sagA* coding region in frame with the mVenus-coding gene (Figure 3. B). Fusions were named TL−S+s, where S and s are the start and stop coordinates (from the start codon), respectively. The fusion TL−109+54 (lacking the first 35 nt) was expressed at levels similar to those of the full-length fusion (TL−144+54). Interestingly, deleting 28 and 43 additional nt (TL−81+54 and TL−66+54) caused a reduction in fluorescence of approximately 50% and 80%, respectively, when compared to the longer fusions (Figure 3. B). However, removing 39 additional nt (TL−27+54) increased the expression to ~40 % of the TL−81+54 truncation (Figure 3. B). These results suggest that the integrity of the region downstream of position −109 of the 5′ UTR is important for *sagA* expression.

In order to test whether the first codons of *sagA* were involved in the repression of the shorter truncations, we removed the CDS on the above-mentioned fusions. Similar to first set, the fusion starting at position −66 (TL−66+3) showed lower expression levels (Figure 3. C). However, all truncations were expressed at higher levels (relative to the longest fusion) than constructs containing the *sagA* CDS fragment. In addition, the shortest fusion (TL−27+3) had similar expression levels that the longest one (Figure 3. C). Together, these results indicate that the absence of the region between −81 and −27 has a negative effect on *sagA* expression, regardless of the presence or absence of the first 18 codons of *sagA* CDS.

To determine whether the repression observed in the shorter truncations was due to inhibition of translation, we constructed a set of transcriptional (TX) fusions containing the same regions of *sagA* 5′ UTR as in the TL fusions. In these fusions, translation of the reporter gene and the fragment of *sagA* CDS are driven by two independent RBSs. Therefore, any effect can be mostly attributed to changes in transcript levels (Figure 3. D). Similar to the results of the translational fusions, constructs starting at positions −81 and −66 of *sagA* 5′ UTR had a lower expression than the ones with the full-length 5′ UTR. However, the differences in the expression of the transcriptional fusions are smaller than in the translational fusions and no difference was observed between the fusions starting at positions −81, −66 and −27. These results indicate that the truncation of 5′ UTR may have an effect on RNA transcription and/or stability.

In order to confirm the importance of the *sagA* 5′ UTR, we tested the expression of a similar set of fusions in *S. pyogenes* using the firefly luciferase gene as a reporter.

In agreement with the results from *E. coli*, the expression of the translational (Figure 3. E) and transcriptional (Figure 3. F) fusions drastically decreased in the fusions lacking 76 nt of the 5′ UTR as compared to the full-length or TX−27+54. Moreover, quantitative reverse-transcription PCR (qRT-PCR) analysis showed that the RNA abundance of TX−66+54 was lower than the full-length and TX−27+54 fusions (Figure 3. G). In contrast to the nearly 90% loss in luminescence (Figure 3. F), RNA abundance of the fusion TX−66+54 was 50% lower than for the full-length fusion (Figure 3. G). Furthermore, RNA levels of TX−27+54 were similar to the full-length fusion (Figure 3. G). The discrepancies between the luciferase activity and the RNA levels might indicate that, at least part of the effect, was due to a reduction of translation. Indeed, even if translation of the reporter gene was driven by a separate RBS, the local ribosome concentration might be affected by the proximity of the *sagA* RBS.

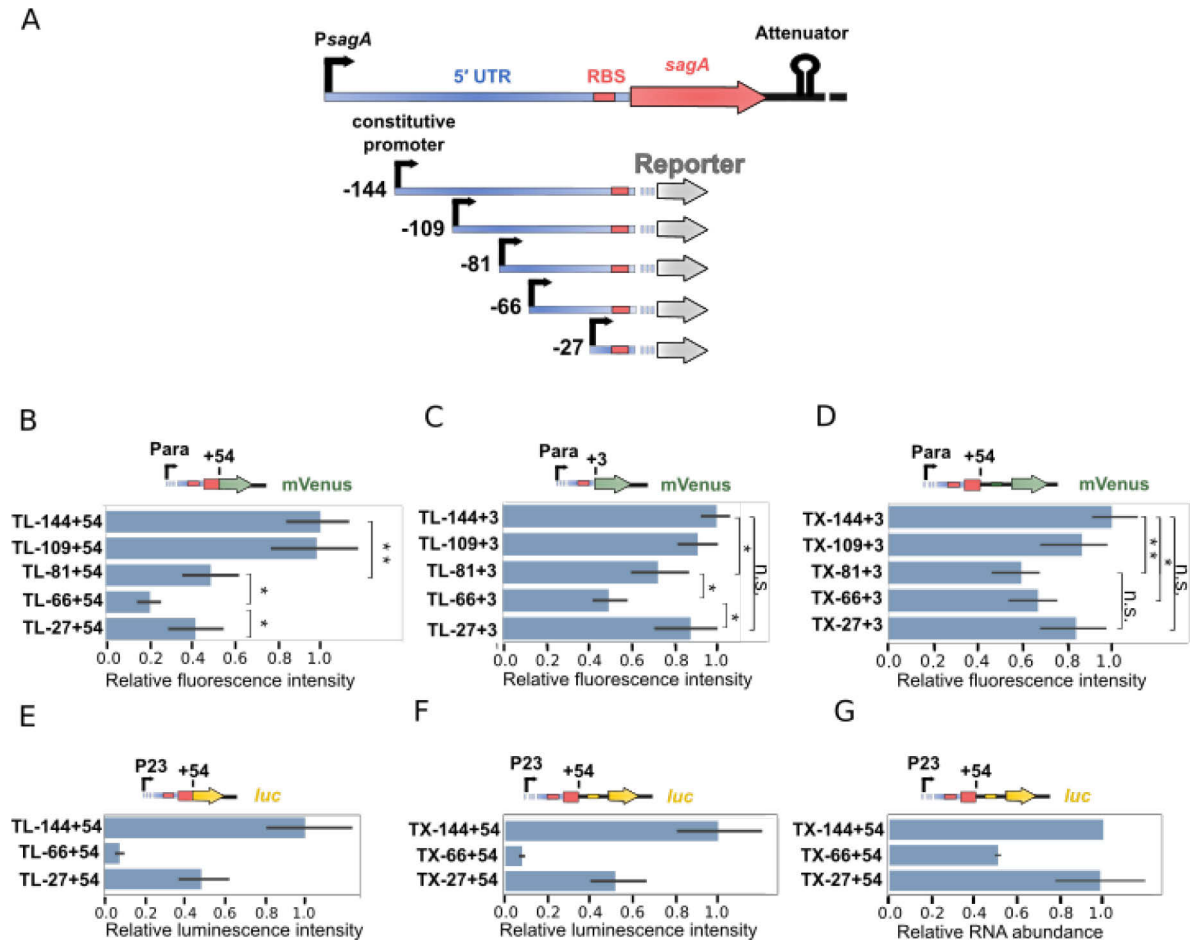

Figure 3. Truncations of *sagA* 5′ UTR affect *sagA* expression levels A) Schematic of reporter fusions containing different fragments of *sagA* 5′ UTR. *sagA* CDS is colored red and the 5′ UTR and RBS are colored blue and green, respectively. B and C) Expression of translational reporter fusions under the arabinose inducible promoter (Para) in *E. coli* containing fragments of *sagA* 5′ UTR with (B) or without (C) the first 54 nt of *sagA* CDS. D) Expression of transcriptional fusions in *E. coli* (similar to B but the translation of mVenus is driven by a separate RBS). Expression of translational (E) or transcriptional (F) reporter constructs containing truncations of *sagA* 5′ UTR fused to the firefly luciferase gene (Luc) under the P23 constitutive promoter in *S. pyogenes.* G) Relative RNA abundance analyzed by qRT−PCR in *S. pyogenes*. In all plots, the labels in the Y−axis indicates the name of the tested fusion. TX and TL are transcriptional and translational fusions, respectively. The numbers indicate the coordinates of the start and end of the *sagA* sequence counting from the *sagA* start codon. Independent t−test *p* values are indicated for relevant comparisons: n.s = $p > 0.05$, * = $p \leq 0.05$, ** = $p \leq 0.01$. The schematic on top of each plot indicates the main features of the tested fusion: fragment of *sagA* CDS (red), reporter gene and configuration (transcriptional or translational). Bars represent the average of at least three biological replicates relative to the empty reporter vector (fluorescence or luminescence intensity) or the longest fusion (qRT− PCR).

Structure changes in truncations might inhibit *sagA* expression

As shown above, the *sagA* 5′ UTR gives rise to an sRNA expressed in a growth phase-dependent manner, suggesting that it is the result of a regulated process. In addition, *sagA* transcript levels are uncoupled from the hemolytic activity (Figure 2. A and B). These results could be explained by changes in RNA structure that regulates translation and/or processing.

Interestingly, structure predictions suggest that the RBS is partially buried in a hairpin (Figure 5. A). In addition, this RBS-blocking structure was predicted to occur with a higher probability in the fusion with impaired expression (starting at position −66) than in fusions showing higher expression levels (Figure 4. B). Furthermore, introducing substitutions that were predicted to stabilize the RBS-blocking structure (Figure 4. A and C) drastically reduced the expression of all fusions, independently of the 5′ UTR fragment that was present (Figure 4. D). However, a set of fusions designed to destabilize the structure, with substitutions in predicted base pairs, (Figure 4. C) failed to abrogate the reduction in expression caused by the truncation of the 5′ UTR (Figure 4. D). This is perhaps due to alternative inhibitory structures formed downstream of the RBS (Figure 4. C).

Analyses of fusions containing additional mutations in the 5′ UTR suggested that the region around position −20 was important for *sagA* expression. Deletion of nucleotides −29 to −20 or −24 to −20, but not deletion of nucleotides −29 to −25, inhibited the luciferase expression to the levels of the fusion that starts at position −27 (Figure 4. A and E). Interestingly, the deletion constructs exhibiting a lower expression were also predicted to have a higher probability of forming an RBS-blocking structure (Figure 4. E). Together, these results suggest that changes in the structure of the 5′ UTR can alter *sagA* expression.

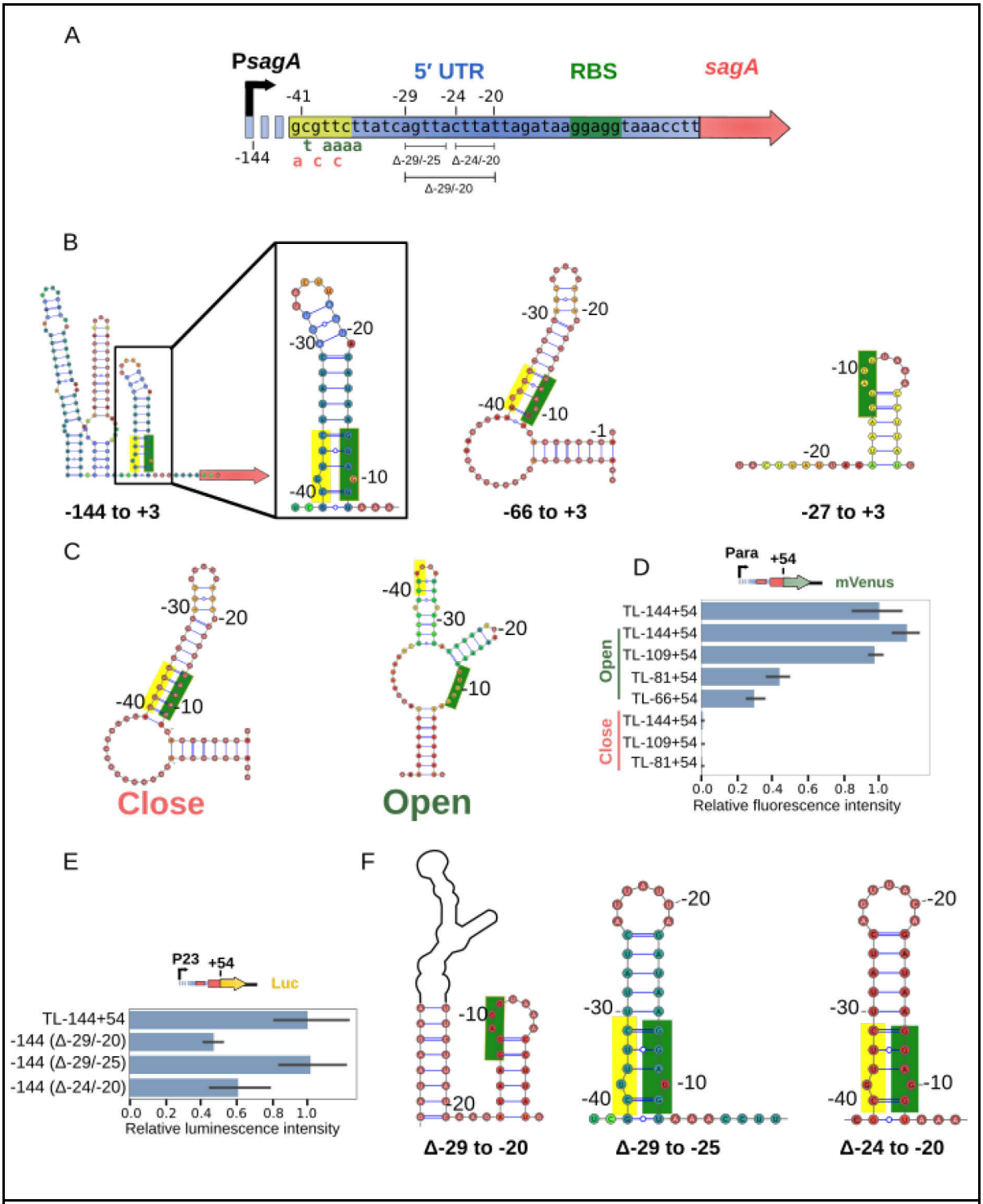

Figure 4. Truncations in *sagA* 5′ UTR affect the structure of the RNA. A) Schematic representation of deletions/substitutions introduced in fusions containing *sagA* 5′ UTR. Numbers indicate coordinates relative to the *sagA* start codon. Green and red letters indicate substitutions introduced to generate the 'open' and 'close' structures, respectively (see panel C). *sagA* CDS is colored red, the RBS and putative anti−RBS are highlighted in green and yellow, respectively. The UTR is colored blue. B) RNA structure prediction of the full−length 5′ UTR (left panel) or truncations (middle and right panels). C) Structure predictions of the 'open' and 'close' mutants, expression of these mutants is shown in D. D) Expression of translational fusions containing different fragments of *sagA* 5′ UTR with substitutions

that are predicted to generate an Open or Close structure (see C). E) Expression of translational fusions containing deletions of *sagA* 5′ UTR as indicated (see panel A). F) Structure prediction of the fusions used in E are shown. All structures were predicted using RNAfold web server (Hofacker et al., 1994) and visualized with varna applet. Color−code indicates the probability associated with the position of each nucleotide (red and blue for high and low probabilities, respectively).

Exposure to metabolite mixes affect the *sagA* 5′ UTR structure

As shown above, structure prediction of the full-length 5′ UTR and truncations indicate that a putative RBS/anti-RBS structure may prevent ribosome binding (Figure 4. A-B). Stabilization/destabilization of this putative RBS-blocking structure might play a role in regulating *sagA* translation, potentially explaining the discrepancy between transcript abundance and hemolytic activity (Figure 2. A and B). The presence of a riboswitch in the *sagA* 5′ UTR could provide a mechanism to regulate accessibility of the RBS and cleavage of the transcript. Indeed, riboswitches control transcript processing and ribosome accessibility, simultaneously (Caron et al., 2012; Shahbabian et al., 2009).

To test the possibility that *sagA* is under the control of a riboswitch, we used a reporter fusion that contained *sagA* 5′ UTR and 18 codons of the CDS fused to the gene of mVenus under an inducible promoter. *E. coli* grown in rich media (EMEM, RPMi or LB) showed a small but reproducible reduction in fluorescence intensity (relative to the vector lacking *sagA* 5′ UTR) compared with *E. coli* grown in minimal media (M9, Figure 5. A). This suggested that *sagA* expression might be inhibited by a molecule present in rich media or a secondary metabolite produced under these conditions. Furthermore, chemical and enzymatic probing of *sagA* 5′ UTR showed that it adopts a stable secondary structure that could act as a riboswitch or another *cis*-regulatory element (Figure 5. B-C).

Next, we aimed to determine whether the structure of the 5′ UTR changes *in vitro* in response to binding of a small molecule. For this, we incubated the *in vitro*-transcribed *sagA* 5′ UTR with different concentrations of yeast extract (YE) and assessed whether specific changes in structure occurred. YE was used as a source of metabolites because most known riboswitches sense molecules that are ubiquitous in nature (Lotz and Suess, 2018) and it has been used before successfully for this purpose (Nelson et al., 2013). In order to detect any changes in the structure of the RNA, we used in-line probing (for a description of the technique see introduction, (Regulski and Breaker, 2008). YE caused a concentration-dependent structure change

on the *sagA* 5′ UTR (Figure 5. D), suggesting that a component in the YE induced a specific conformational change.

It is known that Mg^{2+} is involved in the stabilization of certain RNA structures (Palma et al., 2014). However, a change in structure was still observed upon addition of YE in the absence of Mg^{2+} (Figure 5. E), indicating that Mg^{2+} is not strictly necessary for this YE-dependent conformational change to occur. Additionally, the effect of YE was more evident in absence of Mg^{2+}. For these reasons, successive experiments were carried out in the absence of Mg^{2+}.

One of the proposed functions of SLS-mediated hemolysis is to increase iron availability (Molloy et al., 2011). To determine whether iron was responsible for the observed conformational change, we performed in-line probing in the presence of YE and increasing concentrations of ethylenediaminetetraacetic acid (EDTA), which is a known divalent ion chelator (Figure 5. E). Despite the fact that EDTA inhibited the effect of YE, addition of iron or other divalent cations (in the absence of YE) did not cause specific conformational changes (data not shown, summarized in table 1). Moreover, the effect of YE on the structure of *sagA* 5′ UTR was reduced when the extract was exposed to high temperatures, indicating that the responsible molecule is heat-sensitive (Figure 5. D). This led us to the conclusion that the structural change is not due to the binding of an ion (Figure 5. F).

In order to identify the putative ligand, we performed subsequent (high-performance liquid chromatography) HPLC/in-line probing cycles (in collaboration with the Chemical Biology Department at HZI, Braunschweig). After each in-line probing experiment, the fraction that caused the conformational change was further separated and re-tested (Figure 5. F). The composition and complexity of the fractions was monitored after each cycle by mass spectrometry and promising candidates were selected for individual testing. Interestingly, a reduced number of fractions for each cycle affected the structure of the RNA, suggesting that the structural change was specific and only caused by a limited number of molecules.

Due to the high complexity of the yeast extract, the identification of potential ligands was challenging. In order to simplify the analysis, a less complex and partially characterized metabolite library from *Pseudomonas aeruginosa* was used. Both *P. aeruginosa* cell extract and secreted molecules, in addition to one of the active yeast extract fractions, caused a similar cleavage pattern in a concentration-dependent manner (Figure 5. G). Although several candidates were identified by mass

spectrometry analysis, all tested molecules failed to reproduce the pattern caused by yeast and *P. aeruginosa* extracts (data not shown, see Table 1). This may be due to the fact that the ligand concentration, while still being able to promote RNA structure rearrangements, fell below the limit of detection of mass spectrometry after the HPLC fractionation. A list of the tested compounds can be found in Table 1. These results suggested that the *sagA* 5′ UTR changes confirmation to a distinct structure in response to the presence of a limited number of metabolites. Future experiments should focus on investigating whether these changes translate into regulation of *sagA.*

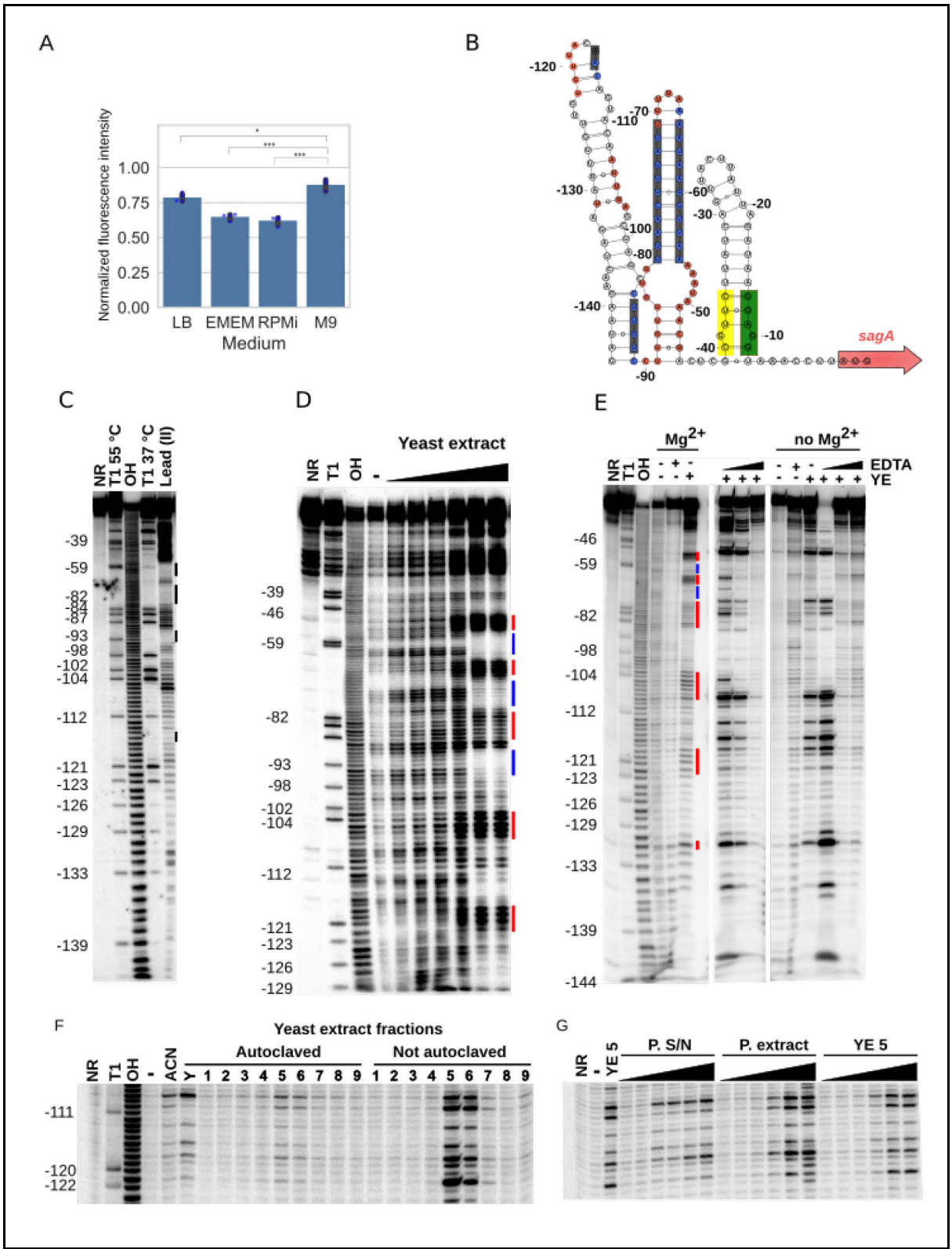

Figure 5. Exposure to metabolites affects the *sagA* 5′ UTR structure. A) Expression of *sagA* 5′ UTR translational fusion to mVenus reporter gene in *E. coli* grown in different media. Fluorescence is normalized against the construct without *sagA* 5′ UTR (pEC2101). Each dot represents an independent biological replicate, bars show the average of these replicates and error bars represent the standard deviation. Independent t−test *p* values are indicated by *: *** = $p \leq 0.001$, * = $p \leq 0.05$.

B) Predicted RNA structure of *sagA* 5′ UTR using RNAfold web server. Numbers indicate coordinates from the start codon. *sagA* CDS is represented by the red arrow. The RBS is highlighted in green. Nucleotides highlighted in red and blue show increased and decreased cleavage by in line probing (see panels D and E), respectively. Grey squares indicate double−stranded regions according to the structure−probing experiment (see panel C). C−G) Structure and in−line probing experiments. Control lanes contain the undigested RNA (no reaction, NR), digested with RNase T1 (T1, that cleaves after every single−stranded G) or alkaline digestion (−OH, resulting in cleavage in every position along the RNA). C) Structure probing of *sagA* 5′ UTR. Treatment of the RNA with Pb^{2+} or RNase T1 that preferentially promotes cleavage of single−stranded regions, protected regions suggest a stable secondary structure (marked with black lines). D and E) Effect of increasing concentrations of yeast extract (YE) on the structure of *sagA* 5′ UTR RNA. Lanes marked with a − show the cleavage pattern in the absence of metabolites. Changes in the cleavage pattern after addition YE indicate that one or more components of metabolite mixes alter RNA structure. Regions that show increased or decreased cleavage rates are marked with red or blue vertical lines, respectively (also indicated in the predicted structure on A). E) Addition of YE affects RNA in the presence or absence of Mg^{2+}. EDTA inhibits the effect of YE on the structure of the RNA. F) Effect of YE fractions produced by reverse phase chromatography (RPC) on RNA structure. Two fractions (5 and 6 containing approximately 50−60% acetonitrile) recapitulate the effect of the complete YE. Autoclaved YE does not induce structure changes. G) *Pseudomonas aeruginosa* secreted metabolites (P. S/N), *P. aeruginosa* cell extract (P. extract) or fraction 5 of RPC (YE5) produce a similar pattern than the complete extract in a concentration−dependent manner. The lane marked as ACN contains 100% acetonitrile, the solvent used for the RPC. RPC was done in collaboration with Franke Raimo and Ulrike Beutling in the Chemical Biology department at the Helmholtz Centre for Infection Research, Braunschweig, Germany.

Table 1: Tested molecules in in-line probing of *sagA* 5′ UTR

Class	Molecules	Concentration (mM)	Effect
Metal Ions	Fe^{+2}, Fe^{+3}, Zn^{+2}, Co^{+2}, Cu^{+2}, Ni^{+2}, Mn^{+2}, Mg^{+2}	1	None or unspecific degradation
Quorum sensing	HHQ, DMQ, PQS	4, 0.4, 0.04	None
Cholate derivatives	Glycocholate, Chenodeoxycholate, Tauroglycocholate, Deoxycholate Methylester, Taurolithocholat−derivative	0.03−0.05	None or unspecific effect

HHQ = 4−hydroxy−2−heptylquinoline

PQS = 2−heptyl−3,4−dihydroxyquinoline

DMQ = 2−Polyprenyl−3−methyl−6−methoxy−1,4−benzoquinone

High-throughput method for riboswitch ligand identification

In-line probing has several limitations as a method to screen for unknown riboswitch ligands. First, the fact that it is labour-intensive and low-throughput limits the number of molecules that can be tested. Second, the changes in RNA structure do not provide information about their biological relevance. Finally, it does not recapitulate kinetic processes that can affect riboswitch activity, such as co-transcriptional folding (Chauvier et al., 2017). Therefore, we aimed to develop a new method that overcomes these limitations allowing us to screen for the putative ligand of *sagA* 5′ UTR and other potential riboswitches.

To this end, we first constructed plasmids that expressed a bicistronic transcript with the mCherry fluorescent protein on the first ORF (open reading frame), and a fusion consisting of the riboswitch of interest in frame with the mVenus-coding gene on the second ORF (Figure 6. A). mCherry served as a control for any riboswitch-independent effect of the metabolites on transcription/translation, while mVenus fluorescence was used to evaluate the activity of the riboswitch. The expression of these fusions was tested in *E. coli* or using *in vitro*-transcription/translation (TL/TX)

reactions exposed to increasing concentrations of the ligand. In order to test the system, we used the predicted thiamine pyrophosphate (TPP) riboswitch from *S. pyogenes* (see below: bioinformatic analysis of riboswitches in *S. pyogenes*). However, no difference in fluorescence was observed in response to addition of TPP either *in vitro* or in *E. coli* (data not shown). The absence of regulation by the TPP riboswitch could be explained by the fact that it was not located at the 5′ end of the transcript, making it unable to adopt a native conformation.

In order to address this issue, the control and test constructs, encoding the mVenus reporter in both cases, were expressed independently under the T7 promoter (Figure 6. B). Nonetheless, increasing concentrations of the ligand failed to modulate reporter gene expression, both *in vitro* and in *E. coli* (data not shown). We hypothesize that expressing the riboswitches using the phage T7 polymerase, which has higher transcription rates than bacterial polymerases, might decrease the sensitivity of certain riboswitches to their ligand. Indeed, transcription speed and pausing has been shown to affect the folding of riboswitches (Chauvier et al., 2017). For this reason, we exchanged the T7 promoter for the bacterial PLac promoter in the constructs (Figure 6. B). In this setting, the expression of the fusion that contains the TPP riboswitch decreased as the concentration of TPP increases (Figure 6. C), while this trend was not observed for the control (Figure 6. D). To account for the unspecific effects of the ligand, we calculated the ratio of the expression between the test and the control (Figure 6. E). However, although a negative trend in the ratio could be observed with increasing concentrations of TPP, the differences were not statistically significant due to the high variability of the results (Figure 6. E). Nonetheless, increasing TPP concentrations correlated with a decrease in the expression of the test but not the control fusions (Figure 6. F-G), indicating that the TPP riboswitch from *S. pyogenes* is able to sense TPP.

The discovery of new riboswitch ligands would potentially involve testing thousands of different molecules. The current method requires testing several ligand concentrations in multiple replicates in order to obtain meaningful correlations. Thus, further optimization will reduce the number of replicates that are needed to evaluate the activity of a riboswitch. Yet, these results indicate that the method developed here can be used for verifying whether a predicted riboswitch is functional or testing a small number of potential ligands.

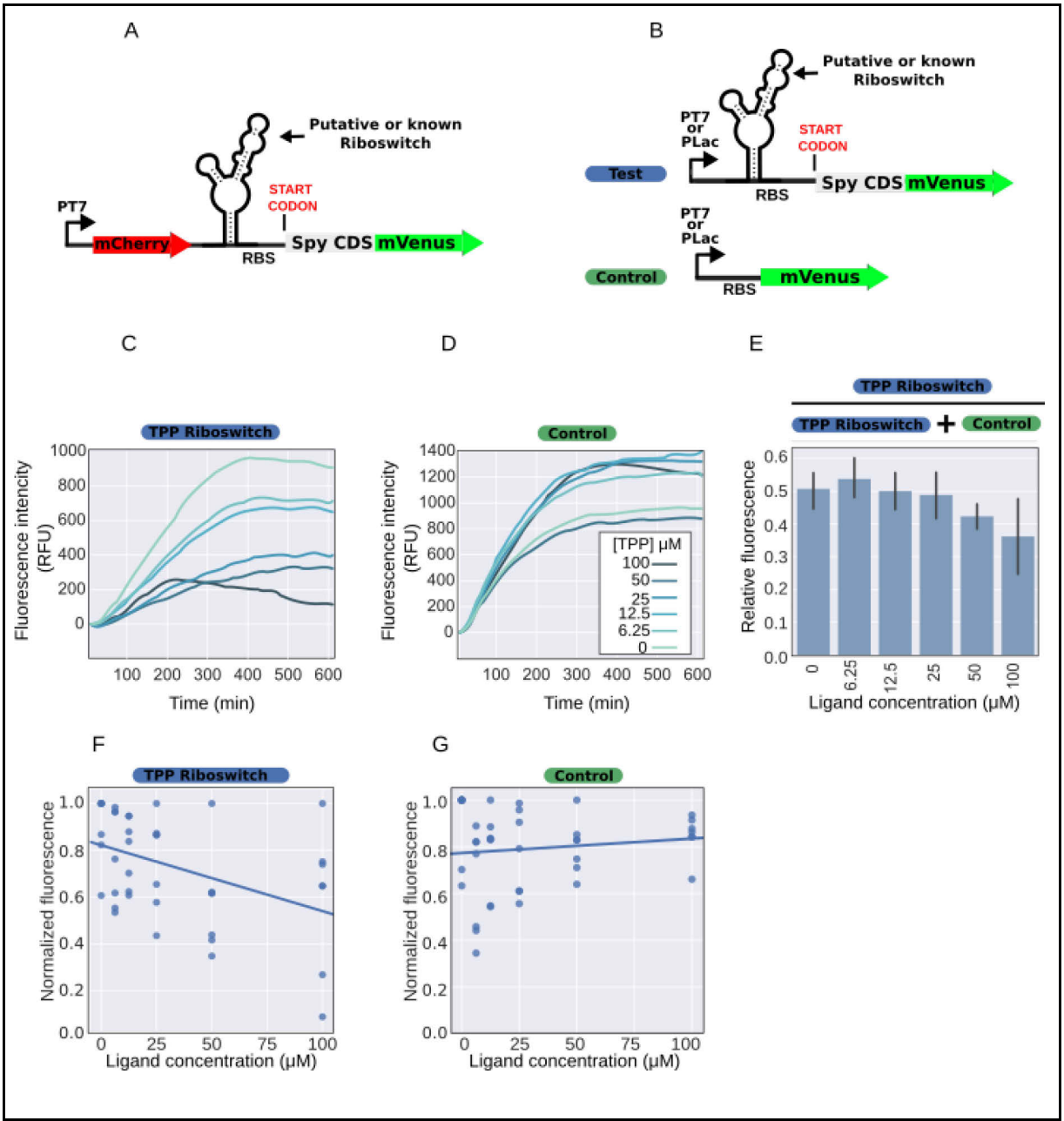

Figure 6. Method for validating riboswitches *in vitro*. A) Reporter construct to test the activity of riboswitches. Bicistronic operon expressed under the T7 promoter encoding the mCherry fluorescent protein as a control (red), followed by a (putative) riboswitch fused to mVenus reporter gene (green). B) Independent monocistronic constructs coding for mVenus (green) under the PLac or T7 promoter. The test constructs (top) encode the putative riboswitch fused to the mVenus gene. The control construct (bottom) encodes only the reporter gene. In all cases, the riboswitches constructs included the first 10 to 18 codons of their native downstream gene fused in frame to the reporter gene. C and D) Representative fluorescence intensity curves over time of an *in vitro* transcription/translation experiment using construct containing the predicted TPP riboswitch of *S. pyogenes* (C) or the control (D). The blue-color gradient indicates the TPP concentration in the reaction from higher (Dark blue, 100 µM) to lower (Light blue, no TPP). E) Normalized maximum fluorescence intensity ratio

(test/(control+test)) of experiments as in (C and D). Bars indicate the average of six individual experiments, error bars represent the standard deviation. F and G) Correlation between ligand concentration and expression of the TPP test (F) or control (F) fusions. The dots represent normalized end-point fluorescence measurements at different TPP concentrations. The line shows a linear regression for these points. Fluorescence of the TPP-test construct in negatively correlated with the concentration of TPP (r=-0.58) whereas the control fusion is not (r=0.23).

Analysis of predicted riboswitches in *S. pyogenes*

The Rfam database (Kalvari et al., 2017, 2018) lists five riboswitches encoded in the genome of *S. pyogenes* (Table 2). To date, only the glycine riboswitch has been experimentally validated (Khani et al., 2018). Therefore, we used the fluorescence *in vitro* TX/TL method to investigate whether the remaining riboswitches respond to their predicted ligand.

Bioinformatics analysis of riboswitches in *S. pyogenes.*

It was shown that single-nucleotide substitutions in the ligand-sensing motif are enough to modify the specificity of riboswitches (Weinberg et al., 2017). Hence, we first investigated whether the predicted riboswitches of *S. pyogenes* deviated from the consensus sequence, which would potentially indicate changes in specificity and/or functionality. To this end, we aligned the sequence of each predicted riboswitch in *S. pyogenes* to the Rfam consensus and identified the deviating nucleotides (Figure 7. A). The high sequence conservation as well as the conservation of base pairs suggested that these predicted riboswitches are functional and sense the predicted ligand (Figure 7. A).

In addition to those predicted by the Rfam database, the presence of two additional riboswitches in *S. pyogenes* was proposed, yybP-ykoY and metk2 (Perez et al., 2009; Rhun et al., 2016). The yybP-ykoY has been reported to sense Mn^{2+} in other organisms and its structure in the bound state has been solved (Price et al., 2015). Although its sequence in *S. pyogenes* deviated significantly from the Rfam consensus, the conservation of nucleotides known to interact with Mn^{2+} (Price et al., 2015) suggests that its function is preserved. The 5′ UTR of the *metk2* gene, which codes for the S-adenosylmethionine (SAM) synthase 2, has been suggested to contain a SAM-sensing riboswitch (Perez et al., 2009). However, sequence and structure prediction analyses suggest that it does not belong to any of the five classes of S-

adenosylmethionine (SAM) riboswitches or of the related metabolite S−adenosylhomocysteine (SAH). Consistent with this, the Rfam database did not predict the presence of a SAM riboswitch in the chromosome of *S. pyogenes* and a sequence-based search did not render any hits. The number of riboswitch classes devoted to sensing SAM suggest that riboswitches have evolved independently more than once to monitor the concentrations of this metabolite. Therefore, it is possible that the 5′ UTR of metk2 belongs to an uncharacterized riboswitch class.

To gain insight into the regulatory mechanism used by these seven riboswitches, we bioinformatically predicted whether transcriptional terminators or RBS-blocking structures are located adjacent to them (Figure 7. B). The predicted mode of regulation and location in the genome of the riboswitches are given in the table below (Table 2).

Table 2. List of predicted riboswitches in *Streptococcus pyogenes*

Riboswitch class	Predicted ligand	Predicted transcriptional terminator	Downstream gene (predicted function)
TPP	thiamine pyrophosphate	yes	*thiT* (thiamine transporter)
FMN	flavin mononucleotide	yes	SPy_0373 (riboflavin transporter)
glycine	glycine	no	SPy_1270 (amino acid symporter)
purine	guanine	no	*xpt* (xanthine phosphoribosyltransferase)
preQ1−II	pre−queuosine	no	SPy_0749 (citrulline cluster−linked gene)
yybP−ykoY*	Mn^{2+}	yes	*pacL* (Ca^{2+} transporter ATPase)
metk2*	S−adenosylmethionine	no	*metk2* (S−adenosylmethionine synthase 2)

Ligand prediction is based on similarity to Rfam consensus. Riboswitches with asterisk (*) are not predicted in Rfam of *S. pyogenes* and ligand is predicted based on genetic context (metk2) or similarity to published structures (yybP-ykoY) (Perez et al., 2009; Le Rhun et al., 2016). Intrinsic transcriptional terminators were predicted using TransTerm (Kingsford et al., 2007).

Functional analysis of predicted riboswitches in *S. pyogenes*

Using the fluorescence *in vitro* TX/TL assay, we confirmed the activity of the TPP, flavin mononucleotide (FMN) and Mn^{2+} riboswitches of *S. pyogenes* (Figure 7. C). As expected, addition of increasing concentrations of the ligand correlated with a reduction in fluorescence in the riboswitch-containing fusions but not in the control construct (Figure 7. C). Surprisingly, the glycine riboswitch failed to respond to its ligand under our experimental conditions. Since the sequence of the glycine aptamer in *S. pyogenes* is conserved (Figure 7. A), it is possible that the riboswitch is not functional under the tested conditions. Indeed, while the concentration used in our assay was in the µM range, a recent publication reported that the glycine concentrations necessary to elicit a response are in the mM range (Khani et al., 2018).

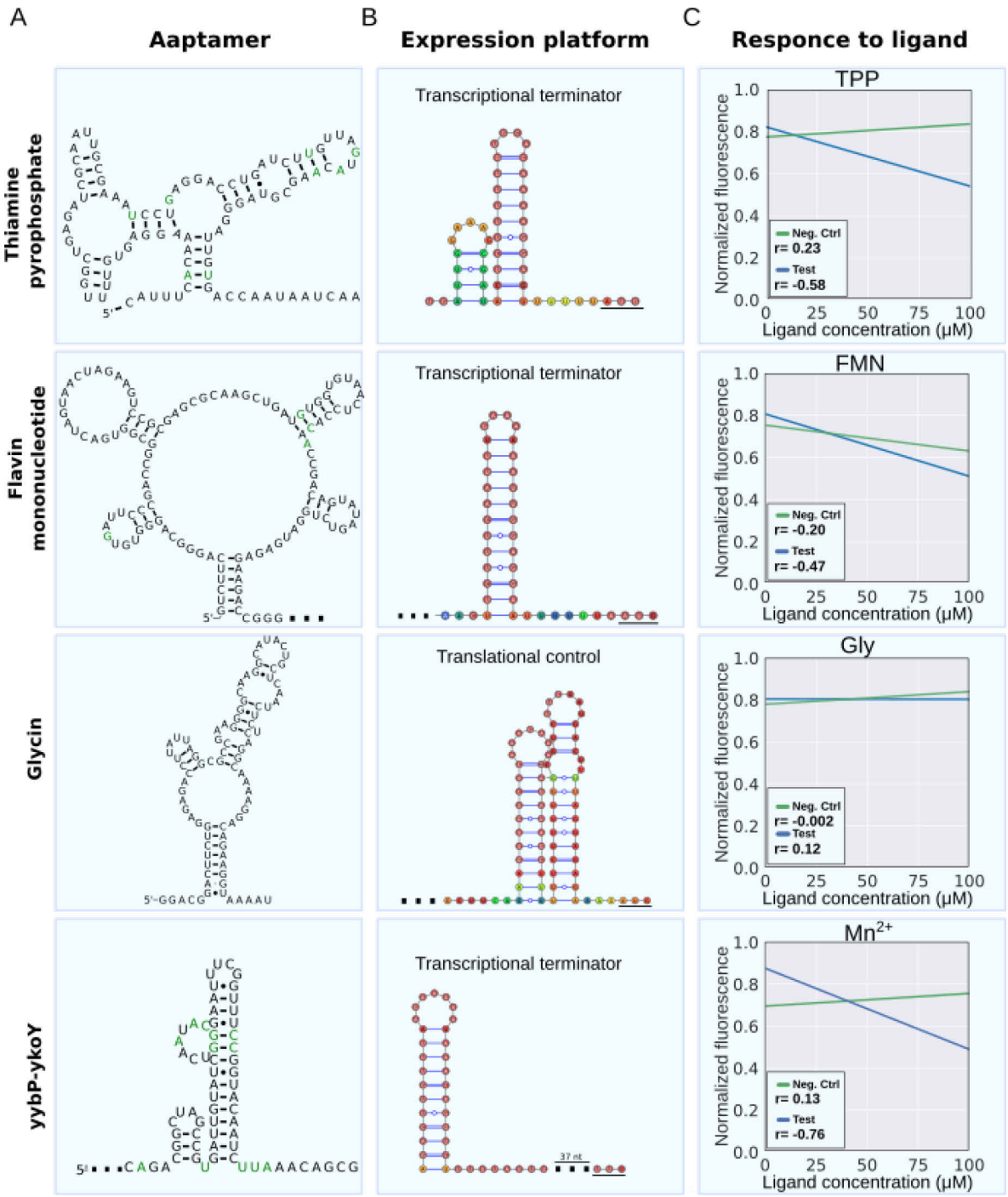

Figure 7. Functional analysis of predicted riboswitches in *S. pyogenes*. A) The aptamer column shows the sequences of selected riboswitches in *S. pyogenes* superimposed with the Rfam consensus structure for each riboswitch class. Nucleotides that deviate from the sequence consensus are colored in green. B) Expression platform column containing the predicted structure (using RNAfold web server) of the expression platform of each riboswitch along with the putative regulatory mechanism (transcriptional or translational depending on the prediction of a transcriptional terminator or an RBS−blocking structure). Start codons are underlined. Color-code indicates the probability associated

with the position of each nucleotide (red and blue for high and low probabilities, respectively). C) Response to ligand showing the correlation between ligand concentration and expression of the test (green) or control (blue) fusions for each riboswitch. Pearson's correlation coefficient (r) is indicated for each correlation. The line shows the linear regression of at least 3 independent replicates for each concentration. The Pearson's correlation coefficient (r) is shown.

Discussion

The aim of this study was to investigate the regulation of *sagA* and its effect in the production of SLS. Here, we report that RNase Y has two independent effects on *sagA*: i) it modulates the production of a sRNA that arises from the 5′ UTR of the *sagA* transcript, ii) it upregulates transcription through an unknown intermediate factor. The mechanism by which RNase Y gives rise to the sRNA remains unknown. It is likely that other factors act together with RNase Y in order to produce a transcript of the observed size (approx. 120 nt). Indeed, if RNase Y directly cleaved at the 3′ end of the sRNA, mutations in this region would abrogate its production. Since this approach has been successful in abrogating RNase Y activity on another target (Broglia et al. 2018), this suggests that RNase Y does not cleave in this region of *sagA* 5′ UTR. In addition, preliminary results suggest that the sRNA is produced from the P23-5′ UTR in the WT and the Δ*rny* strains, albeit the abundance of the sRNA is lower in the latter. The fact that sRNA expression is observed in the Δ*rny* when it is produced from the fusion, but not from the chromosome, can be explained by the higher abundance of the transcript when it is expressed from a plasmid (Figure 1. E), making the sRNA visible even in the production rate is lower.

Three hypotheses could explain these observations (Figure 8). 1) The sRNA is produced by premature transcriptional termination of the *sagA* 5′ UTR in a process that is modulated by RNase Y. 2). A cleavage is produced at the observed 3′ end of the sRNA by an unknown RNase that is regulated by RNase Y and has different requirements than RNase Y, followed by complete degradation of the 3′ fragment. 3) RNase Y cleaves downstream of this region and the sRNA is produced by further trimming by a 3′-5′ exoRNase that stops at this position. These three hypothetical mechanisms would explain the fact that the fragment downstream of the sRNA is undetectable, and that we were unable to inhibit RNase Y cleavage. Further experiments are needed in order to determine whether any of the proposed hypothesis are correct. In the ES growth phase, the lack of detection of the sRNA might indicate

that the sRNA is not produced or that it is unstable. This also suggests that it might be produced by a regulated process. It is currently unclear whether the process that generates the sRNA has any effect on SLS production, as deletion of RNase Y did not have any effect on the expression of the P23-5′ UTR. The high stability of the *sagA* primary transcript and the low abundance of the sRNA suggests the rate at which it is produced is low, explaining the lack of regulation. If the mechanism that generates the sRNA does not give rise to a stable 3′ fragment, the production of the sRNA can only cause a reduction in transcript abundance. On the other hand, if the sRNA is produced by a cleavage in the 5′ UTR that produces a stable 3′ fragment, under certain conditions, this might modulate translation or stability of the downstream fragment. Since we could not detect the 3′ fragment, this latter scenario is unlikely.

Under the studied conditions, the only regulatory activity of RNase Y on *sagA* is transcription induction. Since RNase Y is not able to regulate transcription directly, one or multiple intermediary factors must exist. A transcriptional reporter fusion was constructed to investigate the regions upstream of *sagA* that are involved in RNase Y-mediated regulation. Because the binding sites for some *sagA* regulators are known, this information might narrow down the number of possible regulators. Unexpectedly, the reporter fusion containing the complete intergenic region upstream of *sagA* was not regulated by RNase Y. Introducing the fusion in a plasmid means that there are multiple copies of *sagA* regulatory region per cell. If the number of regulatory molecules is limiting, this would reduce the amount of molecules that are bound to the regulatory region, abrogating the regulation. Introducing the reporter fusions in a low-copy plasmid or in the chromosome will likely address this limitation. Future experiments will focus on the identification of the DNA regions that are essential for *sagA* regulation and the regulator that mediates this effect. Differential expression analysis of Δ*rny* compared to WT will provide a list of potential regulators. The upregulation of *sagA* mRNA by RNase Y is only physiologically relevant if it affects the production of SLS. Interestingly, the hemolysis assay shows that the reduction of *sagA* transcript levels does not automatically entail a lower hemolytic activity, as observed in the ES growth phase. *sagA* transcription is the first of multiple steps in the production of SLS. An increase in the rate of any process from SagA translation to its modification and export, might compensate for the lower transcript abundance. Although more experiments are needed to confirm this hypothesis, there are some pieces of evidence that point to a regulation of *sagA* translation: 1) RNA structure predictions suggest the

presence of an RBS blocking structure (Figure 4A), 2) introducing point-mutations or deletions that are predicted to stabilize this structure inhibits expression of the reporter genes (Figure 4C-F).

We have shown that truncations of *sagA* 5′ UTR have a negative effect in transcript abundance. However, there is evidence to support that at least part of the effect on the production of the reporter proteins comes from translation inhibition. First, the reduction of expression detected by qRT-PCR is lower than the effect on luminescence in *S. pyogenes*. Second, the effect of the truncations is less marked in the transcriptional than in the translational fusions. It remains to be studied whether the reduction in translation affects transcript stability by reducing the number of ribosomes exposing it to RNases. Alternatively, removing the 5′ UTR might affect transcript stability directly by generating an RNA structure that is permissive to cleavage by endoRNases. In addition, removing secondary structures from the 5′ UTR end might allow the 5′-to-3′ exoRNase J1 to degrade the transcript. However, it is unlikely that this accounts for the whole effect given that it is also observed in *E. coli*, which does not seem to encode 5′-to-3′ RNases.

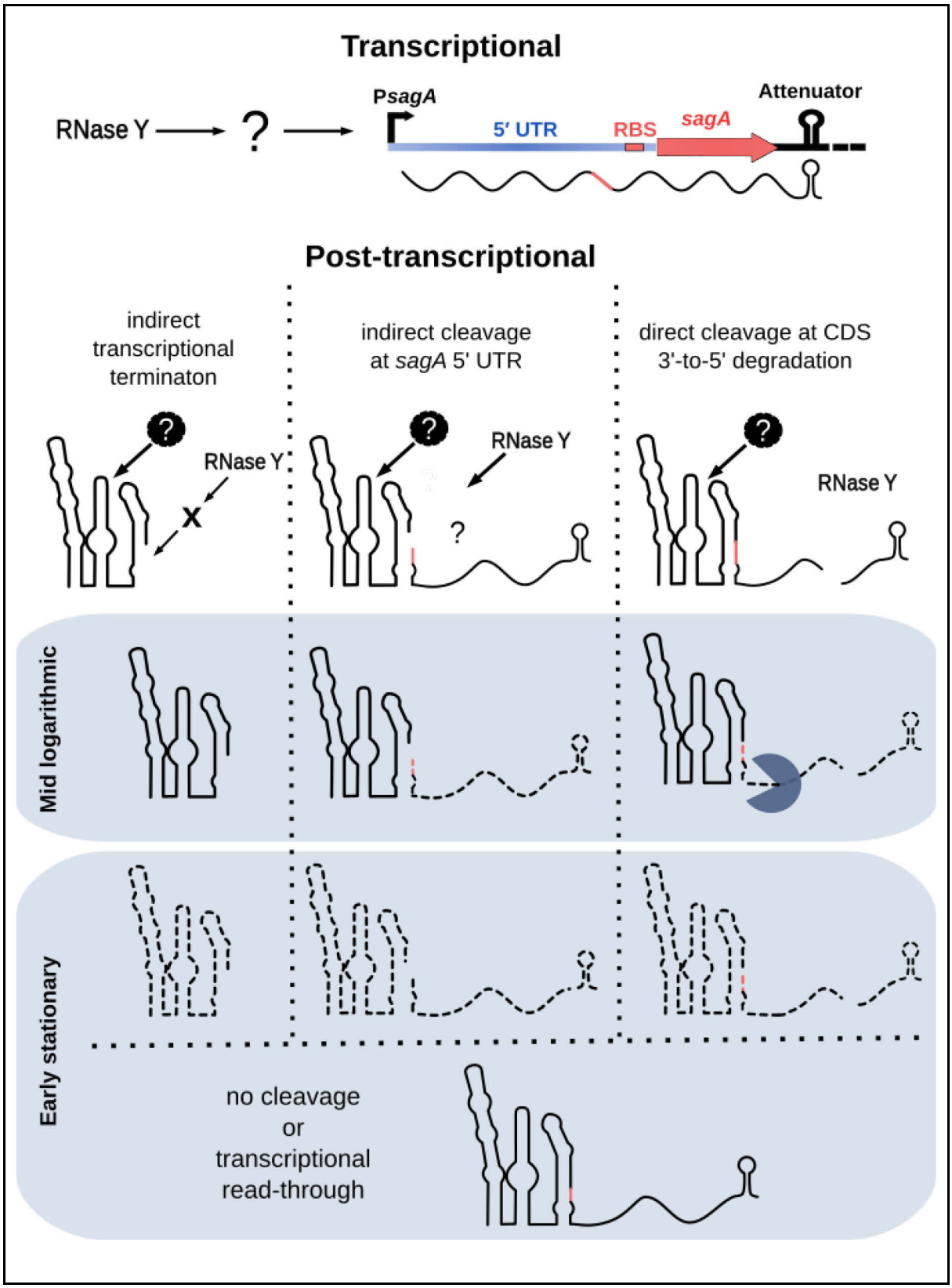
Transcriptional
RNase Y
?
PsagA
5′ UTR
RBS
sagA
Attenuator
Post-transcriptional
indirect transcriptional terminaton
indirect cleavage at sagA 5' UTR
direct cleavage at CDS 3'-to-5' degradation
RNase Y
X
?
Mid logarithmic
Early stationary
no cleavage or transcriptional read-through

Figure 8. Summary of the effects that RNase Y has on *sagA* at the transcriptional and post-transcriptional levels. At the transcriptional level, RNase Y activates P*sagA* indirectly through an unknown factor (represented by the '?') in both mid-logarithmic and early stationary growth phases. At the post-transcriptional level, *sagA* 5′ UTR gives rise to a sRNA that is detectable at mid-logarithmic but not early stationary growth phases. Three hypothetical mechanisms may lead to the production of a sRNA of the detected size. 1) Transcriptional termination modulated by RNase Y indirectly by an unknown factor (represented by the X). 2) RNase Y-regulated cleavage by an unknown RNase (blue scissors) at the 5′ UTR followed by degradation of the downstream fragment. 3) Direct RNase Y (grey scissors) cleavage downstream of the detected 3′-end followed by degradation by a 3′-to-5′ exoRNAse (blue Pacman) until the 3′ end of the sRNA and complete degradation of the downstream fragment. At early stationary growth phase, the sRNA might be unstable or not produced. The *sagA* 5′ UTR might change secondary structure upon binding an unknown ligand (black circle). The *sagA* promoter is represented by the bent arrow. The *sagA* 5′ UTR is coloured blue. The RBS is indicated with a red rectangle. The red arrow represents *sagA* CDS and the hairpin represents the attenuator. Solid or broken undulated lines represent stable or unstable transcripts.

Modulating accessibility to the RBS would allow for a rapid and reversible regulation of *sagA* translation. The reporter fusion experiments in *E. coli* show that the presence of *sagA* 5′ UTR upstream a reporter gene makes the fusion sensitive to changes in growth conditions. This raises the possibility that the 5′ UTR is able to directly sense the presence of a specific molecule and regulate *sagA* translation. We therefore hypothesize that a riboswitch regulates transcription, translation or processing of the transcript. In support of this hypothesis, a synthetic transcript that consists of the *sagA* 5′ UTR changes conformation when exposed to complex metabolite mixes. Several attempts to identify the responsible molecule were unsuccessful. However, the limited number of fractions that were able to induce these changes in structure suggests that the interaction is specific. It has been proposed that molecules, which interact unspecifically with the RNA, appear as a complete or homogeneous degradation pattern on in-line probing experiments (Soukup and Breaker, 1999). Yet, whether the changes in conformation are functionally relevant remains to be seen.

In light of the difficulties of identifying the putative ligand of the *sagA* 5′ UTR, we sought to develop a method that allowed us to screen for ligands while providing functional information. Detection of fluorescent reporter proteins produced by *in vitro* transcription/translation under the control of a riboswitch allowed us to validate the TPP, yybP-ykoY (Mn^{+2}) and FMN riboswitches of *S. pyogenes* for the first time, proving that this method is useful to evaluate the activity of predicted riboswitches.

The high variability of the results obtained during the validation of the assay meant that the experiment has to be performed using multiple ligand concentrations and several replicates to obtain significant correlations. In the current conditions, this method can be used to discriminate among a low number of ligand candidates.

However, further optimization is required in order to use this method to search for a ligand in complex metabolite libraries. In an attempt to reduce the variability, a construct coding mCherry in a separate plasmid was included in each reaction to control for pipetting error. However, this failed to reduce variability across different reactions (data not shown). Increasing reaction volumes or using an *E. coli* extract might help reducing errors in future experiments. However, the high costs of the *in vitro* transcription/translation kits are currently limiting the volume and number of reactions that can be performed.

Materials and Methods

Table 3. List of strains used in the study.

Strain name	**Strain code**	**Relevant characteristics/genotype**	**Source**
Streptococcus pyogenes			
SF370	EC2514	M1 serotype	ATCC® 700294™
SF370	EC2224	M1 serotype	ATCC® 700294™
EC2636		EC2514Δ*sagA*::lox72	Le Rhun unpublished
EC2641		EC2514Δ*sagA5'*UTR::lox72	Le Rhun unpublished
EC2246		EC2224Δ*rny*::lox72	Broglia et al 2018
EC2298		EC2246Δlox72::*rny*-TT3-lox72	Broglia et al 2018
Escherichia coli			
Top10	RDN204	*E. coli* K12 F⁻ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ*ΔM15 Δ*lacX*74 *recA*1 *araD*139 Δ(*ara-leu*)7697 *galU galK* λ⁻ *rpsL*(Str^R) *endA*1 *nupG*	Invitrogen

Table 4. List of plasmids used in the study

Plasmid code	**Relevant characteristics**	**Source**

Plasmids for translational fusions of *sagA* 5'UTR to mVenus (for expression in *E. coli*)		
pBAD18	pBAD18	(Guzman et al., 1995)
pEC2101	pBAD18ΩmVenus	This study
pEC2102	pEC2101Ω5'UTR*sagA*_SagA18codons	This study
pEC2115	pEC2101Ω5'UTR*sagA*-109_+54	This study
pEC2116	pEC2101Ω5'UTR*sagA*-81_+54	This study
pEC2117	pEC2101Ω5'UTR*sagA*-66_+54	This study
pEC2118	pEC2101Ω5'UTR*sagA*-27_+54	This study
pEC2126	pEC2101Ω5'UTR*sagA*-144_+3	This study
pEC2127	pEC2101Ω5'UTR*sagA*-109_+3	This study
pEC2128	pEC2101Ω5'UTRsagA-81_+3	This study
pEC2129	pEC2101Ω5'UTR*sagA*-66_+3	This study
pEC2130	pEC2101Ω5'UTR*sagA*-27_+3	This study
Plasmids for transcriptional fusions of *sagA* 5'UTR to mVenus in *E. coli*		
pEC2133	pEC2101Ω5'UTR*sagA*-144_+54	This study
pEC2134	pEC2101Ω5'UTR*sagA*-109_+54	This study
pEC2135	pEC2101Ω5'UTR*sagA*-81_+54	This study
pEC2136	pEC2101Ω5'UTR*sagA*-66_+54	This study
pEC2137	pEC2101Ω5'UTR*sagA*-27_+54	This study
Plasmid containing *sagA* 5'UTR with substitutions in putative anti-RBS sequences		
pEC2147	pEC2102-40TGAAA	This study
pEC2148	pEC2115-40TGAAA	This study

pEC2149	pEC2116-40TGAAA	This study
pEC2150	pEC2117-40TGAAA	This study
pEC2151	pEC2102-40TACCTC	This study
pEC2152	pEC2115-40TACCTC	This study
pEC2153	pEC2116-40TACCTC	This study
pEC2154	pEC2117-40TACCTC	This study
Plasmids for constructing translational reporter fusions to firefly luciferase in *S. pyogenes*		
pEC2174	pLZ12Km2-P23R:TA:fflucRT	(Loh and Proft, 2013)
pEC2274	pEC2174ΩP*sagA*	This study
pEC2237	pEC2174Ω*sagA*-144+54	This study
pEC2238	pEC2174Ω*sagA*-66+54	This study
pEC2239	PEC2174Ω*sagA*-27+54	This study
pEC2293	pEC2237-Δ10-29	This study
pEC2294	pEC2237-Δ5-29	This study
pEC2295	pEC2237-Δ5-24	This study
pEC2296	pEC2237-ATG-ATT	This study
pEC2297	pEC2237-116G-A	This study
pEC2298	pEC2237-127G-A	This study
pEC2299	pEC2237-132GG-AA	This study
Plasmids for constructing transcriptional reporter fusions to firefly luciferase in *S. pyogenes*		
pEC2270	pEC2174Ω*sagA*-144+54	This study
pEC2271	pEC2174Ω*sagA*-66+54	This study
pEC2272	PEC2174Ω*sagA*-27+54	This study
pEC2273	pEC2173Ω*sagA*-144+54	This study
Plasmids for reporter fusions of putative *S. pyogenes* riboswitches to mVenus (for expression *in vitro*)		
pET21b-RL027	pET21b-RL027	Lentini et al 2013
pEC1985	pET21b-RL027ΩmVenus	This study

pEC2175	pEC1985ΩPlac	This study
pEC2176	pEC2175ΩPlac_SpyTPP+11aa	This study
pEC2177	pEC2175ΩPlac_SpyFMN+14aa	This study
pEC2179	pEC2175ΩPlac_SpyGlycine+13aa	This study
pEC2180	pEC2175ΩPlac_SpyMetk2+10aa	This study
pEC2181	pEC2175ΩPlac_SpyyybP-ykoY+12aa	This study

Table 5. List of primers used in the study.

Primer code	**Target**	**Sequence 5´-3′**	**F/R[a]**
NB[b]			
OLEC3273	5´UTR *sagA* mRNA	AGCTCAATTGTACTGTTGTAACA	R
OLEC7883	*sagA* mRNA	TTATTTACCTGGCGTATAACTTCCGC	R
OLEC288	5S rRNA	AGTTAAGTGACGATAGCCTAG	F
OLEC287	5S rRNA	CTAAGCGACTACCTTATCTCA	R
qRT-PCR[c]			
OLEC8570	fflucRT	TAACCAGTCATTTGCCGCCT	F
OLEC8571	fflucRT	ACGAGCGTGAGAAAAGCGTA	R
qRT-PCR control			
OLEC8574	RepA	TTATTCGCCTTAGGGGAGCG	F
OLEC8575	RepA	CCCCCGTTTCAGCATCAAGA	R
Constructing translational fusions of *sagA* 5'UTR to mVenus in pBAD18			
OLEC3663	XbaI_mVenus	CAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAGCAAAGGCGAAGAACTGTTC	F
OLEC3664	mVenus_BamHI	CAAACTAATTGAATTCGGATCCGCTTATTTATACAGTTCATCCATACCATGCGTAATGCC	R
OLEC7842	*sagA* 5'UTR -144	AAGCGGATCCGAATTCGATAAGAACTAGATAGTTGTTGTGTTACAACAG	F
OLEC7843	*sagA* 5'UTR -109	AAGCGGATCCGAATTCCAATTGAGCTAGCCTTGTCCTTGTTG	F
OLEC7844	*sagA* 5'UTR -81	AAGCGGATCCGAATTCTTAACTTTATTTTTAAAATAAGGTTAAAAATAAACGACTCG	F
OLEC7845	*sagA* 5'UTR -66	AAGCGGATCCGAATTCAAATAAGGTTAAAAATAAACGACTCGCG	F
OLEC7846	*sagA* 5'UTR -27	AAGCGGATCCGAATTCTACTTATTAGATAAGGAGGTAAACCTTATG	F

OLEC7383	*sagA* +54	CGCCTTTGCTCATATGTTGAGTTGTTTCAGCTACACT AGTAGC	R
Constructing translational fusions of *sagA* 5'UTR to mVenus in pBAD18 (with OLEC7842 to 78846)			
OLEC7416	*sagA*	CCTTTGCTCATATGTTATTCTCCTTATAAGTTCAAACT AATGAGTTGTTTCAGCTACACTAGTAGCTAA	R
Introducing substitutions in putative anti-RBS sequences			
OLEC8046	*sagA* 5'UTR	CGACTCGTGAAATTATCAGTTACTTATTAG	F
OLEC8047	*sagA* 5'UTR	CTGATAATTTCACGAGTCGTTTATTTTTAACC	R
OLEC8048	*sagA* 5'UTR	CGACTCACCTCCTTATCAGTTACTTATTAG	F
OLEC8049	*sagA* 5'UTR	CTGATAAGGAGGTGAGTCGTTTATTTTTAACC	R
Constructing translational reporter fusions to firefly luciferase in *S. pyogenes*			
OLEC8380	*sagA* 5'UTR -144	CAGACCTAAGACTGATGACAAAAAGAGAAAATTTTGA TAAAATAGTCTTAGATAAGAACTAGATAGTTGTTGTG TTACAACAGTACAATTG	F
OLEC8381	*sagA* 5'UTR -66	CAGACCTAAGACTGATGACAAAAAGAGAAAATTTTGA TAAAATAGTCTTAAAATAAGGTTAAAAATAAACGACT CGCGTTC	F
OLEC8382	*sagA* 5'UTR -27	CAGACCTAAGACTGATGACAAAAAGAGAAAATTTTGA TAAAATAGTCTTATACTTATTAGATAAGGAGGTAAAC CTTATG	F
OLEC8383	*sagA* 5'UTR +54	GGGGCAGGACCTTTCTTGATATTCTTAGCATCTTCCA TATGTTGAGTTGTTTCAGCTACACTAGTAG	R
Constructing transcriptional reporter fusions to firefly luciferase in *S. pyogenes*			
OLEC7962	*sagA* 5'UTR +54	ACCTGTGAGAATTCTTATTATTGAGTTGTTTCAGCTA CACTAGTAGC	R
OLEC8576	*sagA* up 500	CTAGAGCAGAGCTCCTGTGAAGGTGATGGTAGTTCC ACC	F
OLEC8577	*sagA*	ATCACCATCCGCGGCACTTTTATTATAGTAAAAAATG ATTAATATGTAAACCCTTTC	R
Constructing reporter fusions of putative *S. pyogenes* riboswitches to mVenus (for expression *in vitro*)			
OLEC3774	Spy_FMN	ATAAGCGGATCCGAATTCGTGTCTTCAGGGCAGGGT GTG	F
OLEC3775	Spy_FMN	CGCCTTTGCTCATATGAGAAAGTATACCAATCATAAT CATTTTATGTGTTTTTGACAT	R
OLEC3776	Spy_Gly	ATAAGCGGATCCGAATTCTTAATAAACCGAATGATGT CATGCAGGAGAAG	F
OLEC3777	Spy_Gly	CGCCTTTGCTCATATGCCAAACAAGGTTATCAATTAA TTTAACGAGTGCTATCAT	R
OLEC3782	Spy_TPP	ATAAGCGGATCCGAATTCACTATAATATTTCACAAAG GAGTGCTTTGGCTG	F

OLEC3783	Spy_TPP	CGCCTTTGCTCATATGGATAAGGTATTTGACATTGGTGTTTGGTGAC	R
OLEC3784	Spy_yybP-yko	ATAAGCGGATCCGAATTCTCAAGGGAGTAGCAGACGGCTAG	F
OLEC3785	Spy_yybP-ykoY	CGCCTTTGCTCATATGCGTGTAAAATGCTTCATGTCTTTGTTCTTTAGAC	R

[a] F, forward primer; R, reverse primer
[b] NB, Northern Blot assay usage
[c] qRT-PCR, Quantitative transcription PCR usage

Bacterial strains and growth conditions

Table 3 describes all the bacterial strains used in this study. *E. coli* (Top 10) strain was used as a host for cloning. It was grown at 37 ºC with shaking in Luria Bertani medium. When needed kanamycin antibiotic was added at a final concentration of 25 µg/ml. *S. pyogenes* M1 GAS SF370 (wild type, ATCC 700294) and derivative deletion mutants were cultured at 37 ºC without agitation in a 5% CO2 atmosphere. Todd Hewitt broth (THY) supplemented with 0.2% yeast extract (Servabacter ®) and plates containing tryptic soy agar (TSA) supplemented with 3% sheep blood (Oxoid) were used as liquid and solid media, respectively. When required kanamycin antibiotic was added to the medium at a final concentration of 300 µg/ml. Bacterial growth was monitored by measuring optical density at 620 nm (OD_{620}) with a microplate reader (Eon™, biote ®) using 200 µl of culture. All bacterial strains used in this study were stored at -80ºC. When needed, they were grown over day on a plate and then overnight in 5 ml of THY liquid cultures. For each experiment, 100 ml-flasks containing 50 ml THY were inoculated 1:100 with overnight cultures and grown until the desired OD was reached. In this study, bacteria were collected in two growth phases: ML (OD_{620} = 0.25) and ES (OD_{620} = 0.4).

Bacterial transformation

Plasmids used are listed in Table 4. In order to transform *S. pyogenes,* electrocompetent cells were prepared as in (Caparon and Scott, 1991). Competent cells were stored in 20% ice-cold sterile glycerol at -80 ºC. Bacteria were electroporated in 100 µl of 20% ice-cold sterile glycerol. The OD_{620} of the competent cells was adjusted to 2 or 3 before adding 150 ng and 210 ng of plasmid in the WT and mutant strains (Δ*rny*, Δ*sagA*), respectively. The cells were electroporated in a 0.1 cm electroporation cuvette (Bior) with a pulse of 400 Ω and 25 µF as previously described

by (Perez-Casal et al., 1991) with slight modifications. Immediately after transformation bacteria were transferred to tubes containing 4 ml of THY and incubated for 2 hours. Next, 100 µl of the culture were plated in TSA blood plates supplemented with kanamycin. Plates were incubated for 24 hours and single colonies were used to inoculate 3 ml overnight cultures. Fresh transformations were used for each experiment.

RNA extraction

25 ml of culture was mixed with 25 ml 1:1 acetone:ethanol (prechilled at -20). Total RNA was extracted using TRIzol (Sigma-Aldrich™)/chloroform extraction and isopropanol precipitation from samples collected at ML and ES. RNA concentration and integrity were determined using an UV-spectrophotometer (NanoDrop™, ThermoScientific™) and agarose gel electrophoresis analysis.

Polyacrylamide Northern blot analysis

Northern blot analysis was carried as previously described (Fonfara et al., 2014). Briefly, total RNA was separated on 10% polyacrylamide gels (8 M urea) for approximately 3 h at 100 V and transferred onto nylon membranes (Hybond™ N+, GE healthcare) using Trans-Blot® SD semi-dry transfer apparatus (Biorad) for 45 min at 18 V. The crosslinking was performed using EDC (1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (Sigma) for 1 hour at 60 ºC. Prehybridization was done using Rapid-hyb buffer (GE healthcare) for 15 min at 42°C. The hybridization was carried out overnight at 42 ºC with the previously 5′ radiolabeled oligonucleotide probe (Table 5). The T4-polynucleotide kinase (Fermentas) was used to label 2 µl of (20 pmol/µl) oligonucleotide probes with ^{32}P (0.75 MBq) according to the manufacturer's protocol. The probes were purified with G-25 columns (GE Healthcare). The membranes were then washed with washing buffer I (5X Saline-Sodium Citrate (SSC), 0.1% Sodium dodecyl sulfate (SDS) and then with the washing buffer II (1X SSC, 0.1% SDS) for 15 min. Visualization of the radioactive signal was done using a phosphorimager FLA-9500 (GE HealthCare) after approximately 5 days of exposure. The 30-330-bp AFLP ® DNA Ladder (Invitrogen™) was used as a size marker. The 5S rRNA was used as loading control. The 5S rRNA-specific oligonucleotide probe

(Table 5) was hybridized for 1 h and the membrane was washed and exposed as before.

Rifampicin assay

WT and deletion mutant (Δ*rny* and Δ5′ UTR) strains were grown overnight in 10 ml THY cultures. The overnight cultures were diluted 1:200 in THY medium and grown until ML and ES growth phases. Rifampicin (Sigma-Aldrich) was dissolved in methanol. Once cultures reached the desired OD_{620}, the rifampicin was added at a final concentration of 250 µg/ml. 25 ml of the culture were harvested at the desired time points (0, 30, 60, 90, 120 and 150 min) after rifampicin addition. Afterwards RNA extraction was done as described previously and analyzed by Northern Blot.

Transcriptional luciferase reporter expression

The plasmid-based reporter system (pLZ12Km2-P23R:TA:ffLucRT, Addgene plasmid gift from Thomas Proft) described in (Loh and Proft, 2013) was used to construct plasmid pEC2274, in which the expression of ffluc (firefly luciferase gene) is under the control of the *sagA* promoter region (containing 500 bp upstream the start codon).

Briefly, pLZ12Km2-P23R:TA:ffLucRT was digested with SacI and SacII (Thermo Scientific) to remove the lactococcal constitutive promoter P23. P*sag*A was amplified from WT genomic DNA using primers OLEC8576/OLEC8577 and cloned in pLZ12Km2-P23R:TA:ffLucRT to create plasmid pEC2274. WT and *Δrny* cells were transformed and the activity of *sagA* promoters was measured. After diluting the overnight culture 1:200 in fresh THY, 200 µl of culture per well were added in duplicates in a white opaque 96-well microtiter plate with clear flat bottom (Greiner Bio-One TM) and incubated in the plate reader (synergy, BioTek) at 37º C with 5 % CO_2. Beetle luciferin potassium salt (Promega) was added to each well at a final concentration of 50 ng/µl when the desired OD_{620} was reached (ML for one of the duplicates and ES for the other). Luminescence was measured immediately after luciferin addition using a microplate reader (BioTek™ synergy) with an integration time of 1 sec, with a gain of 120, and a read height of 1 mm. The signal was normalized by dividing it to the signal of the control plasmid (pLZ12Km2-P23R:TA:ffLucRT) and with the luciferase signal obtained from the constitutive promoter P23. The experiments were carried out in independent biological triplicates, each with technical duplicates.

Hemolysis assay

The hemolysis assays were performed as previously described by (Loridan and Alouf, 1986), with modifications. 5 ml defibrinated sheep blood (Oxoid) was washed 3 times with 50 ml cold phosphate buffer saline (PBS). Once the strains of interest reached the desired growth phase needed, 50 ml of bacterial culture were collected and pelleted at 4000 rpm for 15 min at 4 ºC.

Then, bacteria were resuspended in 5 ml PBS for ES and 3 ml PBS for ML. Next, 200 µl were taken and 3 serial dilutions 1:1 were done with PBS in a flat-bottom 96-well plate. After, 50 µl of the bacterial dilutions were transferred to a round-bottom 96-well plate containing 50 µl of the washed blood per well. PBS or Triton 1% was added instead of bacteria as negative and positive controls, respectively. Plates were incubated for 1 h at 37 ºC, and subsequently centrifuged for 15 min at 400 x g at 4 ºC. In order to measure hemoglobin content, 50 µl of the supernatant were collected diluted 1:1 in PBS and transferred in a new 96-well flat-bottom plate to measure absorbance at 540 nm. Serial 1:1 dilutions were made with PBS until values were within readable ranges. To ensure that the hemolysis assay was not saturated, bacterial dilutions where any of the strains reached 100 % lysis (similar to the well with triton) were disregarded. The dilution with the highest bacterial concentration that was not saturated was used for each experiment. The measurements were then normalized against the WT strain for each experiment. Technical duplicates and biological triplicates were performed for each experiment.

qRT-PCR

Quantitative real time PCR (qRT-PCR) experiments were performed using total RNA extracted from the indicated strains. RNA was treated with DNase I enzyme, using TURBO DNA-free™ Kit (Invitrogen), according to the manufacturer's instructions. Reactions were performed in 50 µl containing 50 ng/µl of RNA. The absence of DNA contamination was corroborated by PCR amplification method (using oligos that target the 5S rRNA shown in Table 3). The qRT-PCR was done using the Power SYBR Green RNA- to- CT TM 1 Step Kit (applied biosystems) according to the manufacturer's instructions in 10 µl reactions using primers (shown in Table 5). In order to get the ideal RNA concentration, we performed serial dilutions of the RNA samples adding 1 µl of

RNA (0.01 ng/µl) and we add 0.1 µl of each primer (20 ng/µl) per reaction. The rest of the compounds were added in the volumes indicated by the protocol. While preparing the reactions, we always made common stocks in order to reduce the variability and pipetting errors as much as possible, adding 20% more for each reaction. In order to prove the specificity of the primers used and their respective targets, amplicons melting curve analysis was evaluated. During this assay, we use technical duplicates and the same procedure was performed using biological triplicates.

In vitro transcription

sagA 5′ UTR (as determined by RNAseq) was *in vitro*-transcribed using Ampliscribe T7 Flash transcription kit (epicentre) according to the manufacturer's instructions. The template was produced by PCR using primers OLEC5248 and OLEC5249 and contained the T7 promoter in addition to the 144 nt upstream *sagA* start codon. The PCR product was gel-purified with the Gel Extraction Kit (Qiagen) prior to use. A 40 µl transcription reaction was mixed according to the manufacturer's instructions and incubated at 42 ºC for > 3h. Transcript was treated with DNase I (RNase-free NEB) and ethanol-precipitated. Precipitated RNA was gel-purified by electrophoresis on a 10 % polyacrylamide gel with 4 M urea. The gel was stained with ethidium bromide, bands were cut and RNA eluted at 50 ºC with eluRNA solution (0.3 M Na acetate, 0.5 EDTA and 0.1% SDS). Eluted RNA was ethanol-precipitated and stored at -20 ºC.

Labelling and purification

Before labelling, RNA was dephosphorylated using FastAP™ (Fermentas). Briefly, a 20 µl reaction containing 1 µM RNA and 1 µl FastAP was incubated at 37 ºC for at least 15 min. RNA was purified using RNA clean and concentrator-5 (Zymo Research). After, 30 pmol of dephosphorylated RNA was 5′-end-labelled with P^{32} using PNK (Fermentas) for 30 min at 37 ºC. Next, RNA was purified using Illustra MicroSpin G-25 columns (GE Healthcare), gel-purified and precipitated as before. Finally, RNA was resuspended in 60 µl of water and stored at – 20 ºC until used.

RNA structure probing

To determine the structure of *sagA* 5′ UTR, the *in vitro* transcribed RNA was treated with either RNase T1 (Ambion), RNase III (Ambion) or lead(II) acetate (Sigma Aldrich).

For structure-sensitive RNase T1 or RNase III digestion, 10 µl reactions were prepared (according to the manufacturer's instructions) containing 1 µ labelled RNA, 1 µl 10X structure buffer, 1 µg Yeast RNA, 1 µl RNase. Reaction was incubated for 3 min at 37 ºC and reaction was stopped by adding 10 µl of 2x urea loading buffer (10 M urea, 1.5 mM EDTA, pH 8). Lead (II) was carried out by mixing 1 µl labelled RNA with 1 µl structure buffer, 1 ug yeast RNA, 1 µl Lead acetate 25 mM (freshly prepared) and 6 µl H_2O. Reaction was incubated for 1 min at 37 ºC and stopped with 10 µl urea loading buffer. RNase T1 was also used to generate a ladder under denaturing (structure-insensitive) conditions by incubating 5 min at 55 ºC the following reaction: 1 µl 10X Buffer (0.25 M Na citrate pH 5.0), 7 µl urea loading buffer and 1 µl labelled RNA. The reaction was stopped with 3 µl loading buffer and 7 µl H2O. Alkaline (OH) ladder was generated by incubating for 10 min at 90 ºC a 10 µl reaction containing: 1 µl labelled RNA and 1 µl Na_2CO_3 10X (0.5 M Na_2CO_3 pH 9.0, 10 mM EDTA).

In-line probing

In-line probing experiments were done as previously described (Regulski and Breaker, 2008). RNA was incubated overnight at 37 ºC in 10 µl reactions containing 1 µl labelled RNA and 5 µl 2x reaction buffer (100 mM Tris-HCl pH 8.3, 200 mM KCl with or without 40 mM $MgCl_2$ as indicated), ions or yeast extract was dissolved in water and added to the reaction in the indicated concentrations.

After incubation reaction, 10 µl urea loading buffer was added and samples were resolved by electrophoresis in sequencing 10 % polyacrylamide gel with 8 M urea. The gels were finally exposed overnight in a BAS Storage Phosphor Screens (GE HealthCare) and developed using a Typhoon FLA 9500 (GE Healthcare).

Reverse phase chromatography

Yeast extract powder (servabacter) was dissolved in water to 100 mg/ml and fractionated by reverse phase chromatography using a Octadecyl Solid Phase Extraction Column (JT Baker).

Fluorescent *in vitro* transcription/translation assay

In vitro transcription/translation reactions were carried out using the PurExpress kit (NEB). Reactions were done in a final volume of 6 µl containing 1.7 µl Solution A, 1.25

µl Solution B, 0.04 µl RNasin, 1 µl *E. coli* RNA polymerase holoenzyme (NEB), 1 µl ligand (6x) and 1 µl template plasmid (Table 4). Reaction mixes were prepared for the test and control plasmids separately, without the ligand. Then 5 µl of the mix were added to each well of a black plate with 148 low volume wells with transparent flat bottom.

Finally, 1 µl of diluted 6x ligand was added to the mix and fluorescence was read in a plate reader. The reaction was carried out at 37 ºC for 10 h and measurements were taken every 30 min. Because in these conditions, protein amount cannot decrease, any reduction in fluorescence was taken as an artefact. Therefore, the maximum fluorescence value for each reaction was taken for further analysis. Then, the values from the different experiments were normalized against the highest value for each construct. The expression ratio between the test and the control was calculated by dividing the normalized value for the test construct by the sum of the test and the control. These normalized values were also used to obtain the correlations between the ligand concentration and the expression of the construct. The linear regression was plotted using python seaborn.lmplot library. The Pearson's correlation coefficient (r) is shown for each regression and was calculated using scipy.stats.pearsonr python library.

Contributions

Anaïs Le Rhun did the the RNAseq analysis and preliminary Northern blot analyses of *sagA* 5′ UTR in WT and Δ*rny* strains. Anne-Laure Lécrivain constructed the *rny* deletion and complemented strains and analyzed the expression profile of RNase Y in different growth phases. Victoria Gabriel contributed to the cloning of constructs and performed the Northern blots and expression analysis of *sagA reporters.* HPLC was performed by Raimo Franke and Ulrike Beutling from the Chemical Biology Department at the Helmholtz Centre for Infection Research, Braunschweig.

References

Agarwal, S., Agarwal, S., Pancholi, P., and Pancholi, V. (2011). Role of Serine/Threonine Phosphatase (SP-STP) in Streptococcus pyogenes Physiology and Virulence. J. Biol. Chem. 286, 41368–41380.

Ames, T.D., and Breaker, R.R. (2010). Bacterial Riboswitch Discovery and Analysis. In The Chemical Biology of Nucleic Acids, Günteryer, ed. (John Wiley & Sons, Ltd), pp. 433–454.

Barnett, T.C., Bugrysheva, J.V., and Scott, J.R. (2007). Role of mRNA Stability in Growth Phase Regulation of Gene Expression in the Group A Streptococcus. J. Bacteriol. *189*, 1866–1873.

Barrick, J.E., and Breaker, R.R. (2007). The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol. *8*, R239.

Barrick, J.E., Corbino, K.A., Winkler, W.C., Nahvi, A., Mandal, M., Collins, J., Lee, M., Roth, A., Sudarsan, N., Jona, I., et al. (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc. Natl. Acad. Sci. U. S. A. *101*, 6421–6426.

Baruch, M., Belotserkovsky, I., Hertzog, B.B., Ravins, M., Dov, E., McIver, K.S., Le Breton, Y.S., Zhou, Y., Youting, C.C., and Hanski, E. (2014). An Extracellular Bacterial Pathogen Modulates Host Metabolism to Regulate its Own Sensing and Proliferation. Cell *156*, 97–108.

Bastiat-Sempe, B., Love, J.F., Lomayesva, N., and Wessels, M.R. (2014). Streptolysin O and NAD-Glycohydrolase Prevent Phagolysosome Acidification and Promote Group A Streptococcus Survival in Macrophages. MBio *5*, e01690-14.

Betschel, S.D., Borgia, S.M., Barg, N.L., Low, D.E., and Azavedo, J.C.S.D. (1998). Reduced Virulence of Group A Streptococcal Tn916 Mutants That Do Not Produce Streptolysin S. Infect. Immun. *66*, 1671–1679.

Breaker, R.R. (2011). Prospects for Riboswitch Discovery and Analysis. Mol. Cell *43*, 867–879.

Breaker, R.R. (2012). Riboswitches and the RNA World. Cold Spring Harb. Perspect. Biol. *4*, a003566.

Broglia, L., Materne, S., Lécrivain, A.-L., Hahnke, K., Le Rhun, A., and Charpentier, E. (2018). RNase Y-mediated regulation of the streptococcal pyrogenic exotoxin B. RNA Biol. 1–12.

Bugrysheva, J.V., and Scott, J.R. (2010). The ribonucleases J1 and J2 are essential for growth and have independent roles in mRNA decay in Streptococcus pyogenes. Mol. Microbiol. *75*, 731–743.

Caron, M.-P., Bastet, L., Lussier, A., Simoneau-Roy, M., Massé, E., and Lafontaine, D.A. (2012). Dual-acting riboswitch control of translation initiation and mRNA decay. Proc. Natl. Acad. Sci. U. S. A. *109*, E3444-3453.

Chauvier, A., Picard-Jean, F., Berger-Dancause, J.-C., Bastet, L., Naghdi, M.R., Dubé, A., Turcotte, P., Perreault, J., and Lafontaine, D.A. (2017). Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. Nat. Commun. *8*, 13892.

Chen, Z., Itzek, A., Malke, H., Ferretti, J.J., and Kreth, J. (2013). Multiple Roles of RNase Y in Streptococcus pyogenes mRNA Processing and Degradation. J. Bacteriol. *195*, 2585–2594.

Chiarot, E., Faralla, C., Chiappini, N., Tuscano, G., Falugi, F., Gambellini, G., Taddei, A., Capo, S., Cartocci, E., Veggi, D., et al. (2013). Targeted Amino Acid Substitutions Impair Streptolysin O Toxicity and Group A Streptococcus Virulence. MBio *4*, e00387-12.

Cunningham, M.W. (2016). Post-Streptococcal Autoimmune Sequelae: Rheumatic Fever and Beyond. In Streptococcus Pyogenes : Basic Biology to Clinical Manifestations, J.J. Ferretti, D.L. Stevens, and V.A. Fischetti, eds. (Oklahoma City (OK): University of Oklahoma Health Sciences Center), p.

Dalton, T.L., and Scott, J.R. (2004). CovS inactivates CovR and is required for growth under conditions

of general stress in Streptococcus pyogenes. J. Bacteriol. *186*, 3928–3937.

Datta, V., Myskowski, S.M., Kwinn, L.A., Chiem, D.N., Varki, N., Kansal, R.G., Kotb, M., and Nizet, V. (2005). Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection. Mol. Microbiol. *56*, 681–695.

DebRoy, S., Gebbie, M., Ramesh, A., Goodson, J.R., Cruz, M.R., Hoof, A. van, Winkler, W.C., and Garsin, D.A. (2014). A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. Science *345*, 937–940.

DebRoy, S., Saldaña, M., Travisany, D., Montano, A., Galloway-Peña, J., Horstmann, N., Yao, H., González, M., Maass, A., Latorre, M., et al. (2016). A Multi-Serotype Approach Clarifies the Catabolite Control Protein A Regulon in the Major Human Pathogen Group A *Streptococcus*. Sci. Rep. *6*, 32442.

DeLoughery, A., Lalanne, J.-B., Losick, R., and Li, G.-W. (2018). Maturation of polycistronic mRNAs by the endoribonuclease RNase Y and its associated Y-complex in Bacillus subtilis. Proc. Natl. Acad. Sci. U. S. A.

Deutscher, J., Francke, C., and Postma, P.W. (2006). How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol. Mol. Biol. Rev. MMBR *70*, 939–1031.

Durand, S., and Condon, C. (2018). RNases and Helicases in Gram-Positive Bacteria. Microbiol. Spectr. *6*.

Engleberg, N.C., Heath, A., Vardaman, K., and DiRita, V.J. (2004). Contribution of CsrR-Regulated Virulence Factors to the Progress and Outcome of Murine Skin Infections by Streptococcus pyogenes. Infect. Immun. *72*, 623–628.

Feng, W., Minor, D., Liu, M., and Lei, B. (2017). Requirement and Synergistic Contribution of Platelet-Activating Factor Acetylhydrolase Sse and Streptolysin S to Inhibition of Neutrophil Recruitment and Systemic Infection by Hypervirulent emm3 Group A Streptococcus in Subcutaneous Infection of Mice. Infect. Immun. *85*, e00530-17.

Fontaine, M.C., Lee, J.J., and Kehoe, M.A. (2003). Combined Contributions of Streptolysin O and Streptolysin S to Virulence of Serotype M5 Streptococcus pyogenes Strain Manfredo. Infect. Immun. *71*, 3857–3865.

Gao, J., Gusa, A.A., Scott, J.R., and Churchward, G. (2005). Binding of the Global Response Regulator Protein CovR to the sag Promoter of Streptococcus pyogenes Reveals a New Mode of CovR-DNA Interaction. J. Biol. Chem. *280*, 38948–38956.

Georg, J., and Hess, W.R. (2018). Widespread Antisense Transcription in Prokaryotes. Microbiol. Spectr. *6*.

Goldmann, O., Sastalla, I., Wos-Oxley, M., Rohde, M., and Medina, E. (2009). Streptococcus pyogenes induces oncosis in macrophages through the activation of an inflammatory programmed cell death pathway. Cell. Microbiol. *11*, 138–155.

Gordon, C. (2007). The two faces of Janus: virulence gene regulation by CovR/S in group A streptococci. Mol. Microbiol. *64*, 34–41.

Graham, M.R., Virtaneva, K., Porcella, S.F., Barry, W.T., Gowen, B.B., Johnson, C.R., Wright, F.A., and Musser, J.M. (2005). Group A Streptococcus transcriptome dynamics during growth in human blood reveals bacterial adaptive and survival strategies. Am. J. Pathol. *166*, 455–465.

Gryllos, I., Levin, J.C., and Wessels, M.R. (2003). The CsrR/CsrS two-component system of group A Streptococcus responds to environmental Mg2+. Proc. Natl. Acad. Sci. *100*, 4227–4232.

Gryllos, I., Tran-Winkler, H.J., Cheng, M.-F., Chung, H., Bolcome, R., Lu, W., Lehrer, R.I., and Wessels, M.R. (2008). Induction of group A Streptococcus virulence by a human antimicrobial peptide. Proc. Natl. Acad. Sci. U. S. A. *105*, 16755–16760.

Harder, J., Franchi, L., Muñoz-Planillo, R., Park, J.-H., Reimer, T., and Núñez, G. (2009). Activation of the Nlrp3 Inflammasome by Streptococcus pyogenes Requires Streptolysin O and NF-κB Activation but Proceeds Independently of TLR Signaling and P2X7 Receptor. J. Immunol. Baltim. Md 1950 *183*, 5823–5829.

Hidalgo-Grass, C., Ravins, M., Dan-Goor, M., Jaffe, J., Moses, A.E., and Hanski, E. (2002). A locus of group A Streptococcus involved in invasive disease and DNA transfer. Mol. Microbiol. *46*, 87–99.

Higashi, D.L., Biais, N., Donahue, D.L., Mayfield, J.A., Tessier, C.R., Rodriguez, K., Ashfeld, B.L., Luchetti, J., Ploplis, V.A., Castellino, F.J., et al. (2016). Activation of band 3 mediates group A *Streptococcus* streptolysin S-based beta-haemolysis. Nat. Microbiol. *1*, nmicrobiol20154.

Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., and Schuster, P. (1994). Fast folding and comparison of RNA secondary structures. Monatshefte Für Chem. Chem. Mon. *125*, 167–188.

Hollands, K., Proshkin, S., Sklyarova, S., Epshtein, V., Mironov, A., Nudler, E., and Groisman, E.A. (2012). Riboswitch control of Rho-dependent transcription termination. Proc. Natl. Acad. Sci. U. S. A. *109*, 5376–5381.

Horstmann, N., Saldaña, M., Sahasrabhojane, P., Yao, H., Su, X., Thompson, E., Koller, A., and Iii, S.A.S. (2014). Dual-Site Phosphorylation of the Control of Virulence Regulator Impacts Group A Streptococcal Global Gene Expression and Pathogenesis. PLOS Pathog. *10*, e1004088.

Hynes, W., and Sloan, M. (2016). Secreted Extracellular Virulence Factors. In Streptococcus Pyogenes : Basic Biology to Clinical Manifestations, J.J. Ferretti, D.L. Stevens, and V.A. Fischetti, eds. (Oklahoma City (OK): University of Oklahoma Health Sciences Center), p.

Ignatov, D., and Johansson, J. (2017). RNA-mediated signal perception in pathogenic bacteria. Wiley Interdiscip. Rev. RNA.

Jimenez, J.C., and Federle, M.J. (2014). Quorum sensing in group A Streptococcus. Front. Cell. Infect. Microbiol. *4*.

Kaito, C., Kurokawa, K., Matsumoto, Y., Terao, Y., Kawabata, S., Hamada, S., and Sekimizu, K. (2005). Silkworm pathogenic bacteria infection model for identification of novel virulence genes. Mol. Microbiol. *56*, 934–944.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2017). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. *46*, D335–D342.

Kang, S.O., Caparon, M.G., and Cho, K.H. (2010). Virulence Gene Regulation by CvfA, a Putative RNase: the CvfA-Enolase Complex in Streptococcus pyogenes Links Nutritional Stress, Growth-Phase Control, and Virulence Gene Expression. Infect. Immun. *78*, 2754–2767.

Khani, A., Popp, N., Kreikemeyer, B., and Patenge, N. (2018). A glycine riboswitch in Streptococcus pyogenes controls expression of a sodium:alanine symporter family protein gene. Front. Microbiol. *9*.

Khemici, V., Prados, J., Linder, P., and Redder, P. (2015). Decay-Initiating Endoribonucleolytic Cleavage by RNase Y Is Kept under Tight Control via Sequence Preference and Sub-cellular Localisation. PLoS Genet *11*, e1005577.

Kietzman, C.C., and Caparon, M.G. (2010). CcpA and LacD.1 Affect Temporal Regulation of Streptococcus pyogenes Virulence Genes. Infect. Immun. *78*, 241–252.

Kinkel, T.L., and McIver, K.S. (2008). CcpA-Mediated Repression of Streptolysin S Expression and Virulence in the Group A Streptococcus. Infect. Immun. *76*, 3451–3463.

Laalami, S., Bessières, P., Rocca, A., Zig, L., Nicolas, P., and Putzer, H. (2013). Bacillus subtilis RNase Y Activity In Vivo analyzed by Tiling Microarrays. PLOS ONE *8*, e54062.

Lee, E.R., Baker, J.L., Weinberg, Z., Sudarsan, N., and Breaker, R.R. (2010). An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. Science *329*, 845–848.

Lee, S.W., Mitchell, D.A., Markley, A.L., Hensler, M.E., Gonzalez, D., Wohlrab, A., Dorrestein, P.C., Nizet, V., and Dixon, J.E. (2008). Discovery of a widely distributed toxin biosynthetic gene cluster. Proc. Natl. Acad. Sci. *105*, 5879–5884.

Limbago, B., Penumalli, V., Weinrick, B., and Scott, J.R. (2000). Role of Streptolysin O in a Mouse Model of Invasive Group A Streptococcal Disease. Infect. Immun. *68*, 6384–6390.

Lin, A., Loughman, J.A., Zinselmeyer, B.H., Miller, M.J., and Caparon, M.G. (2009). Streptolysin S Inhibits Neutrophil Recruitment during the Early Stages of Streptococcus pyogenes Infection. Infect. Immun. *77*, 5190–5201.

Loh, E., Dussurget, O., Gripenland, J., Vaitkevicius, K., Tiensuu, T., Mandin, P., Repoila, F., Buchrieser, C., Cossart, P., and Johansson, J. (2009). A trans-Acting Riboswitch Controls Expression of the Virulence Regulator PrfA in Listeria monocytogenes. Cell *139*, 770–779.

Lotz, T.S., and Suess, B. (2018). Small-Molecule-Binding Riboswitches. Microbiol. Spectr. *6*.

Mandal, M., and Breaker, R.R. (2004). Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat. Struct. Mol. Biol. *11*, 29–35.

Mangold, M., Siller, M., Roppenser, B., Vlaminckx, B.J.M., Penfound, T.A., Klein, R., Novak, R., Novick, R.P., and Charpentier, E. (2004). Synthesis of group A streptococcal virulence factors is controlled by a regulatory RNA molecule. Mol. Microbiol. *53*, 1515–1527.

Marincola, G., and Wolz, C. (2017). Downstream element determines RNase Y cleavage of the saePQRS operon in Staphylococcus aureus. Nucleic Acids Res. *45*, 5980–5994.

Marincola, G., Schäfer, T., Behler, J., Bernhardt, J., Ohlsen, K., Goerke, C., and Wolz, C. (2012). RNase Y of Staphylococcus aureus and its role in the activation of virulence genes. Mol. Microbiol. *85*, 817–832.

Maxson, T., Deane, C.D., Molloy, E.M., Cox, C.L., Markley, A.L., Lee, S.W., and Mitchell, D.A. (2015). HIV protease inhibitors block streptolysin S production. ACS Chem. Biol. *10*, 1217–1226.

Mehdizadeh Aghdam, E., Hejazi, M.S., and Barzegar, A. (2016). Riboswitches: From living biosensors to novel targets of antibiotics. Gene *592*, 244–259.

Mellin, J.R., Tiensuu, T., Bécavin, C., Gouin, E., Johansson, J., and Cossart, P. (2013). A riboswitch-regulated antisense RNA in Listeria monocytogenes. Proc. Natl. Acad. Sci. *110*, 13132–13137.

Mellin, J.R., Koutero, M., Dar, D., Nahori, M.-A., Sorek, R., and Cossart, P. (2014). Sequestration of a two-component response regulator by a riboswitch-regulated noncoding RNA. Science *345*, 940–943.

Meyer, M.M., Hammond, M.C., Salinas, Y., Roth, A., Sudarsan, N., and Breaker, R.R. (2011). Challenges of ligand identification for riboswitch candidates. RNA Biol. *8*, 5–10.

Miyoshi-Akiyama, T., Takamatsu, D., Koyanagi, M., Zhao, J., Imanishi, K., and Uchiyama, T. (2005). Cytocidal effect of Streptococcus pyogenes on mouse neutrophils in vivo and the critical role of streptolysin S. J. Infect. Dis. *192*, 107–116.

Molloy, E.M., Cotter, P.D., Hill, C., Mitchell, D.A., and Ross, R.P. (2011). Streptolysin S-like virulence factors: the continuing sagA. Nat. Rev. Microbiol. *9*, 670–681.

Mozola, C.C., and Caparon, M.G. (2015). Dual modes of membrane binding direct pore formation by Streptolysin O. Mol. Microbiol. *97*, 1036–1050.

Nagata, M., Kaito, C., and Sekimizu, K. (2008). Phosphodiesterase Activity of CvfA Is Required for

Virulence in Staphylococcus aureus. J. Biol. Chem. *283*, 2176–2184.

Nelson, J.W., Sudarsan, N., Furukawa, K., Weinberg, Z., Wang, J.X., and Breaker, R.R. (2013). Riboswitches in eubacteria sense the second messenger c-di-AMP. Nat. Chem. Biol. *9*, 834–839.

Nilsson, M., Sørensen, O.E., Mörgelin, M., Weineisen, M., Sjöbring, U., and Herwald, H. (2006). Activation of human polymorphonuclear neutrophils by streptolysin O from Streptococcus pyogenes leads to the release of proinflammatory mediators. Thromb. Haemost. *95*, 982–990.

Nizet, V., Beall, B., Bast, D.J., Datta, V., Kilburn, L., Low, D.E., and Azavedo, J.C.S.D. (2000). Genetic Locus for Streptolysin S Production by Group A Streptococcus. Infect. Immun. *68*, 4245–4254.

Obana, N., Nakamura, K., and Nomura, N. (2017). Role of RNase Y in Clostridium perfringens mRNA Decay and Processing. J. Bacteriol. *199*.

Perez, N., Treviño, J., Liu, Z., Ho, S.C.M., Babitzke, P., and Sumby, P. (2009). A Genome-Wide Analysis of Small Regulatory RNAs in the Human Pathogen Group A Streptococcus. PLoS ONE *4*, e7668.

Price, I.R., Gaballa, A., Ding, F., Helmann, J.D., and Ke, A. (2015). Mn2+-Sensing Mechanisms of yybP-ykoY Orphan Riboswitches. Mol. Cell *57*, 1110–1123.

Regulski, E.E., and Breaker, R.R. (2008). In-line probing analysis of riboswitches. Methods Mol. Biol. Clifton NJ *419*, 53–67.

Rhun, A.L., Beer, Y.Y., Reimegård, J., Chylinski, K., and Charpentier, E. (2016). RNA sequencing uncovers antisense RNAs and novel small RNAs in Streptococcus pyogenes. RNA Biol. *13*, 177–195.

Salim, K.Y., Azavedo, J.C. de, Bast, D.J., and Cvitkovitch, D.G. (2007). Role for sagA and siaA in Quorum Sensing and Iron Regulation in Streptococcus pyogenes. Infect. Immun. *75*, 5011–5017.

Salim, K.Y., de Azavedo, J.C., Bast, D.J., and Cvitkovitch, D.G. (2008). Regulation of sagA, siaA and scpC by SilCR, a putative signaling peptide of Streptococcus pyogenes. FEMS Microbiol. Lett. *289*, 119–125.

Saroj, S.D., Maudsdotter, L., Tavares, R., and Jonsson, A.-B. (2016). Lactobacilli Interfere with Streptococcus pyogenes Hemolytic Activity and Adherence to Host Epithelial Cells. Front. Microbiol. *7*.

Saroj, S.D., Holmer, L., Berengueras, J.M., and Jonsson, A.-B. (2017). Inhibitory role of acyl homoserine lactones in hemolytic activity and viability of Streptococcus pyogenes M6 S165. Sci. Rep. *7*, 44902.

Sesto, N., Wurtzel, O., Archambaud, C., Sorek, R., and Cossart, P. (2013). The excludon: a new concept in bacterial antisense RNA-mediated gene regulation. Nat. Rev. Microbiol. *11*, 75–82.

Shahbabian, K., Jamalli, A., Zig, L., and Putzer, H. (2009). RNase Y, a novel endoribonuclease, initiates riboswitch turnover in Bacillus subtilis. EMBO J. *28*, 3523–3533.

Shelburne, S.A., Keith, D., Horstmann, N., Sumby, P., Davenport, M.T., Graviss, E.A., Brennan, R.G., and Musser, J.M. (2008). A direct link between carbohydrate utilization and virulence in the major human pathogen group A Streptococcus. Proc. Natl. Acad. Sci. *105*, 1698–1703.

Shelburne, S.A., Olsen, R.J., Suber, B., Sahasrabhojane, P., Sumby, P., Brennan, R.G., and Musser, J.M. (2010). A combination of independent transcriptional regulators shapes bacterial virulence gene expression during infection. PLoS Pathog. *6*, e1000817.

Shewell, L.K., Harvey, R.M., Higgins, M.A., Day, C.J., Hartley-Tassell, L.E., Chen, A.Y., Gillen, C.M., James, D.B.A., Alonzo, F., Torres, V.J., et al. (2014). The cholesterol-dependent cytolysins pneumolysin and streptolysin O require binding to red blood cell glycans for hemolytic activity. Proc. Natl. Acad. Sci. *111*, E5312–E5320.

Sierig, G., Cywes, C., Wessels, M.R., and Ashbaugh, C.D. (2003). Cytotoxic Effects of Streptolysin O and Streptolysin S Enhance the Virulence of Poorly Encapsulated Group A Streptococci. Infect. Immun. *71*, 446–455.

Soukup, G.A., and Breaker, R.R. (1999). Relationship between internucleotide linkage geometry and the stability of RNA. RNA *5*, 1308–1325.

Steiner, K., and Malke, H. (2001). relA-Independent Amino Acid Starvation Response Network of Streptococcus pyogenes. J. Bacteriol. *183*, 7354–7364.

Stevens, D.L., and Bryant, A.E. (2016). Severe Group A Streptococcal Infections. In Streptococcus Pyogenes : Basic Biology to Clinical Manifestations, J.J. Ferretti, D.L. Stevens, and V.A. Fischetti, eds. (Oklahoma City (OK): University of Oklahoma Health Sciences Center), p.

Storz, G., Vogel, J., and Wassarman, K.M. (2011). Regulation by small RNAs in bacteria: expanding frontiers. Mol. Cell *43*, 880–891.

Sudarsan, N., Lee, E.R., Weinberg, Z., Moy, R.H., Kim, J.N., Link, K.H., and Breaker, R.R. (2008). Riboswitches in Eubacteria Sense the Second Messenger Cyclic Di-GMP. Science *321*, 411–413.

Sumitomo, T., Nakata, M., Higashino, M., Jin, Y., Terao, Y., Fujinaga, Y., and Kawabata, S. (2011). Streptolysin S Contributes to Group A Streptococcal Translocation across an Epithelial Barrier. J. Biol. Chem. *286*, 2750–2761.

Sundar, G.S., Islam, E., Braza, R.D., Silver, A.B., Le Breton, Y., and McIver, K.S. (2018). Route of Glucose Uptake in the Group a Streptococcus Impacts SLS-Mediated Hemolysis and Survival in Human Blood. Front. Cell. Infect. Microbiol. *8*.

Timmer, A.M., Timmer, J.C., Pence, M.A., Hsu, L.-C., Ghochani, M., Frey, T.G., Karin, M., Salvesen, G.S., and Nizet, V. (2009). Streptolysin O Promotes Group A Streptococcus Immune Evasion by Accelerated Macrophage Apoptosis. J. Biol. Chem. *284*, 862–871.

Vajjala, A., Biswas, D., Tay, W.H., Hanski, E., and Kline, K.A. (2018). Streptolysin-induced endoplasmic reticulum stress promotes group A Streptococcal host-associated biofilm formation and necrotising fasciitis. Cell. Microbiol. E12956.

Valdes, K.M., Sundar, G.S., Belew, A.T., Islam, E., El-Sayed, N.M., Breton, Y.L., and McIver, K.S. (2018). Glucose Levels Alter the Mga Virulence Regulon in the Group A Streptococcus. Sci. Rep. *8*, 4971.

Vega, L.A., Malke, H., and McIver, K.S. (2016). Virulence-Related Transcriptional Regulators of Streptococcus pyogenes. In Streptococcus Pyogenes : Basic Biology to Clinical Manifestations, J.J. Ferretti, D.L. Stevens, and V.A. Fischetti, eds. (Oklahoma City (OK): University of Oklahoma Health Sciences Center), p.

Waters, L.S., and Storz, G. (2009). Regulatory RNAs in bacteria. Cell *136*, 615–628.

Watson, M.E., Neely, M.N., and Caparon, M.G. (2016). Animal Models of Streptococcus pyogenes Infection. In Streptococcus Pyogenes : Basic Biology to Clinical Manifestations, J.J. Ferretti, D.L. Stevens, and V.A. Fischetti, eds. (Oklahoma City (OK): University of Oklahoma Health Sciences Center), p.

Weinberg, Z., Barrick, J.E., Yao, Z., Roth, A., Kim, J.N., Gore, J., Wang, J.X., Lee, E.R., Block, K.F., Sudarsan, N., et al. (2007). Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. Nucleic Acids Res. *35*, 4809–4819.

Weinberg, Z., Nelson, J.W., Lünse, C.E., Sherlock, M.E., and Breaker, R.R. (2017). Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. Proc. Natl. Acad. Sci. U. S. A.

Wessels, M.R. (2016). Pharyngitis and Scarlet Fever. In Streptococcus Pyogenes : Basic Biology to Clinical Manifestations, J.J. Ferretti, D.L. Stevens, and V.A. Fischetti, eds. (Oklahoma City (OK): University of Oklahoma Health Sciences Center), p.

Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A., and Breaker, R.R. (2004). Control of gene expression by a natural metabolite-responsive ribozyme. Nature *428*, 281–286.

Yoshino, M., Murayama, S.Y., Sunaoshi, K., Wajima, T., Takahashi, M., Masaki, J., Kurokawa, I., and Ubukata, K. (2010). Nonhemolytic Streptococcus pyogenes isolates that lack large regions of the sag operon mediating streptolysin S production. J. Clin. Microbiol. *48*, 635–638.

Chapter Two:

Regulatory roles of Cas9 in *Francisella novicida*

Introduction

As mentioned in chapter one, a specific class of ncRNAs are part of the CRISPR adaptive immune systems. These are present in approximately 87% of archaeal and 50% of bacterial sequenced genomes (Jackson et al., 2017) and defend prokaryotic cells against foreign nucleic acids. CRISPR-Cas systems are commonly composed of CRISPR RNAs encoded by the CRISPR array and the CRISPR-associated (Cas) proteins. The array contains a variable number of identical repeats interspaced with unique sequences, known as spacers. The number and identity of the Cas proteins varies across different CRISPR-Cas types.

Despite the diversity of CRISPR-Cas systems, they all achieve immunity through three general stages: 1) Acquisition, also known as adaptation, 2) expression/processing and 3) interference (Hille et al., 2018).

Acquisition occurs when a bacteriophage or a plasmid invades a bacterial cell, and part of its genetic material is integrated in the CRISPR array, generating a new spacer. Cas1 and Cas2, two Cas proteins found in most CRISPR-Cas systems, are involved in spacer acquisition (Makarova et al., 2011, 2015; Nuñez et al., 2014, 2015a, 2015b).

This new spacer constitutes a “memory” device that allows identification and targeting of the same threat upon reinfection (see below). The expression/processing stage consists of transcription of the array, to produce a precursor CRISPR RNA (pre-crRNA), followed by maturation, which involves the specific cleavage of the pre-crRNA to produce various crRNAs. Each mature crRNA is comprised of a repeat (or part of it) and a spacer (or part of it). In the final stage, interference, an invading genetic molecule containing a complementary sequence to the spacer (protospacer), is recognized and digested, preventing its maintenance and replication.

The array was first discovered in 1987 (Ishino et al., 1987) and in 2002 its association with *cas* genes was first noticed (Jansen et al., 2002). However, the function of the array and the Cas proteins remained obscure. In 2005, the homology of the spacers to phages and other mobile genetic elements led to the hypothesis that CRISPR-Cas constitutes a defense system, which was confirmed in 2007 (Barrangou

et al., 2007; Bolotin et al., 2005; Makarova et al., 2006; Mojica et al., 2005). Soon after, it was shown that in some types of CRISPR-Cas (types I and III), the crRNAs guide a complex of Cas proteins to destroy phages by targeting DNA (Brouns et al., 2008; Marraffini and Sontheimer, 2008).

To date, multiple RNA and DNA-targeting CRISPR-Cas systems have been discovered and classified in two classes and six types (Koonin et al., 2017; Makarova et al., 2013; Shmakov et al., 2015, 2017). Class 1 includes types I, III and IV, which use a complex of multiple Cas proteins to carry out interference. In contrast, class 2 systems use only one Cas protein in the interference step and is comprised of types II, V and VI.

Because of their simplicity class 2 systems, especially the type II systems, have been used for genetic engineering (see below).

Type II CRISPR-Cas systems

Type II systems, further subdivided in three subtypes (A-C), are DNA-targeting systems characterized by the presence of Cas9 and the *trans*-activating crRNA (tracrRNA) (Chylinski et al., 2013; Deltcheva et al., 2011). Type II-A systems contain the characteristic *csn2* gene (Makarova et al., 2015; Shmakov et al., 2017, Koonin et al., 2017), which codes for a protein that is involved in spacer acquisition (Heler et al., 2015; Ka et al., 2018; Wei et al., 2015). In contrast, type II-B systems lack *csn2* but encode Cas4, which is also involved in acquisition (Kieper et al., 2018; Lee et al., 2018; Shiimori et al., 2018). Type II-C contain only the genes coding for Cas9, Cas1 and Cas2. The length and sequence of Cas9 vary from one subtype to another (Chylinski et al., 2013).

tracrRNA contains an anti-repeat sequence that mediates the formation of a duplex with the pre-crRNA repeats (Deltcheva et al., 2011). The duplex is promoted and stabilized by Cas9. Once the Cas9:tracrRNA:crRNA complex is formed, the duplex of RNAs is co-processed by RNase III. Further trimming of the RNAs by unknown RNases gives rise to the mature tracrRNA:crRNA duplex, which remains bound to Cas9 (Deltcheva et al., 2011). In the type II-C systems of *Neisseria meningitidis* and *Campylobacter jejuni*, each crRNA is transcribed from their own promoter, located within each repeat (Dugar et al., 2013; Zhang et al., 2013). Though RNase III processing of the duplex is still observed, this is not required for activity (Zhang et al., 2013).

Binding of Cas9 to tracrRNA:crRNA triggers a conformational change that renders Cas9 capable of searching for target protospacers that are complementary to the spacer of crRNA (Jinek et al., 2014). In order to avoid targeting the CRISPR array in the bacterial chromosome, Cas9 only checks for complementarity in the sequence that is next to a PAM. Because the PAM is present next to the protospacer but absent in the CRISPR array, this prevents self-targeting by Cas9. The sequence of the PAM varies depending of the Cas9 orthologue (e.g. NGG for *S. pyogenes* Cas9) (Gasiunas et al., 2012; Jinek et al., 2012). Upon DNA invasion, Cas9 in complex with the tracrRNA:crRNA duplex samples the available PAMs and starts unwinding the DNA helix upstream of the PAM in search of a protospacer. Both, the contact with the PAM and the spacer-protospacer base-pairing are essential requirements to activate Cas9 DNA endonuclease activity, which leads to a double-strand break of its target (Jinek et al., 2014; Sternberg et al., 2014). Although some mismatches can be tolerated between the spacer and the protospacer, a PAM-proximal seed sequence of 10- to 12-nt must be fully complementary for Cas9 to cleave (Anders et al., 2014; Jiang et al., 2013; Jinek et al., 2012; Sternberg et al., 2014; Szczelkun et al., 2014). However, 9 base pairs in the PAM-proximal region are enough for stable binding of Cas9 to the DNA (Singh et al., 2016).

Cas9 structure and biochemistry

Cas9 is a multi-domain protein that contains three conserved features: two nuclease domains (HNH and RuvC-like) and an Arg-rich motif. The HNH and the RuvC-like domains cleave the complementary and non-complementary strands in the protospacer, respectively, producing a double-strand break. Both nuclease domains require Mg^{2+} to cleave DNA (Jinek et al., 2012).

Cas9 is arranged in two lobes, named the recognition (REC) lobe and the nuclease (NUC) lobe. The REC lobe consists of three regions: the bridge helix, the REC1 domain and REC2 domains. The NUC lobe contains the RuvC, the HNH and the PAM-interacting domain. This latter one determines the PAM specificity. The RNA:target DNA complex is enclosed between the REC and the NUC lobes. The Arg-rich motif, located in the bridge helix, is shown to interact with the spacer (Nishimasu et al., 2014). This interaction appears to be sequence-independent, in contrast to the repeat:anti-repeat region of the duplex, recognized by the REC lobe, which appears to be sequence-specific. Mismatches between the repeat and the anti-repeat that do not

alter the structure of the duplex are tolerated (Briner et al., 2014).

While the arginine residues on the bridge helix are highly conserved in all type II CRISPR-Cas systems, the length and sequence of the REC lobe varies among the different Cas9 orthologs. This explains the impossibility of Cas9 to function with tracrRNA:crRNA duplexes from divergent CRISPR-Cas systems (Fonfara et al., 2013).

Biotechnological applications of Cas9

The discovery that Cas9 can be guided to cleave any sequence that is next to the PAM (Jinek et al., 2012) quickly led to the development of genome-editing tools that are easily programmed to target virtually any gene of interest in almost any organism including human cells (Barrangou and Doudna, 2016; Jinek et al., 2012; Cong et al., 2013; Doudna and Charpentier, 2014; Jinek et al., 2013; Makarova et al., 2011; Mali et al., 2013). Once Cas9 cleaves the desired sequence, the cell attempts to repair the DSB by non-homologous end-joining (NHEJ) or homology-directed repair (HDR). NHEJ is error-prone and may lead to gene knockout by introducing frameshift mutations. HDR uses a donor that bares homology to the break-flanking regions and, if provided exogenously, allows introducing a desired sequence in the target gene.

The first modification to the natural system was fusing the tracrRNA:crRNA duplex to generate a single-guide (sg)RNA that can be transcribed as one molecule (Jinek et al., 2012). Then, inactivation of one or both nuclease domains allowed turning Cas9 into a nickase (n)Cas9 or a dead (d)Cas9, which is catalytically inactivated. nCas9 may be used to reduce the risk of cleaving undesired sequences (off-targets). By directing two nCas9 molecules to sequences that are in close proximity, a DSB is only generated when the two molecules cleave the desired sequence, reducing the risk of off-targets (Ran FA., et al 2013).

dCas9 can be used to repress transcription by targeting promoter regions. In addition, it can serve to guide effector proteins (such as transcriptional activators or methylases) to the region of interest by fusing them to dCas9 (Dominguez et al., 2016; Hilton et al., 2015; Kearns et al., 2015). The flexibility and programmability of this system means that new applications are constantly emerging.

Non-canonical roles of the type II CRISPR-Cas systems

Artificial modification of CRISPR components have generated systems with diverse functionalities, however some CRISPR-Cas systems have (naturally) evolved

functions beyond immunity (Louwen et al., 2014; Ratner et al., 2015; Westra et al., 2014). Because type II systems are mostly found in pathogenic and commensal bacteria, (Chylinski et al., 2013; Fonfara et al., 2013; Sampson et al., 2013), the regulatory functions of these systems, and their role in virulence regulation have been studied. For example, *cas9* deletion mutants in *Streptococcus agalactiae*, *C. jejuni*, *N. meningitidis*, and *Francisella novicida* are attenuated in virulence, attachment to or intracellular survival in host cells (Louwen et al., 2013; Ma et al., 2018; Sampson et al., 2013). The type II-A *S. agalactiae* Cas9 (SagCas9) is involved in adherence to host cells and survival to phagocytosis by macrophages (Ma et al., 2018). Consequently, a strain deleted for *cas9* is attenuated in virulence in murine and zebrafish infection models (Ma et al., 2018). Furthermore, SagCas9 was shown to repress the expression of the transcriptional regulator *regR,* which in turn inhibits the expression of the virulence factor hyaluronidase (Ma et al., 2018). SagCas9 was proposed to mediate degradation of *regR* mRNA due to partial complementarity with the CRISPR array (Ma et al., 2018). However, how SagCas9 promotes transcript degradation remains unknown. The type II-C *Campylobacter jejuni* Cas9 (CjCas9) is also involved in adhesion, invasion, translocation and cytotoxicity as observed by *in vitro* infection of cells (Louwen et al., 2013). Interestingly, a recent study has shown that CjCas9 can cleave RNAs that are complementary to the crRNA spacer in a tracrRNA-dependent manner (Dugar et al., 2018). RNA targeting by CjCas9 is PAM-independent and mediated by its HNH nuclease domain (Dugar et al., 2018). Although this raises the possibility that CjCas9 regulates gene expression by cleaving RNA, a direct link between RNA targeting and regulation of expression remains to be established. Similarly, *Neisseria meningitidis* Cas9 (NmeCas9), that is also type II-C, has been involved in virulence regulation (Sampson et al., 2013) and has a PAM-independent and RNA-mediated RNase activity *in vitro* (Rousseau et al., 2018) but the link between the RNase activity and gene regulation has not been established.

A role in virulence has also been proposed for the CRISPR-Cas type II-B system in *F. novicida*. *F. novicida* is an intracellular pathogen used as a model to study the highly infectious and extremely virulent *F. tularensis*, which is the causative agent of tularemia and a potential bioweapon (Kingry and Petersen, 2014). Once *F. novicida* is phagocytized by the macrophages, it escapes the phagosome and replicates in the cytosol. At least two receptors can recognize intracellular *F. novicida*, the Toll-like receptor 2 (TLR2) and the AIM2/ASC inflammasome (Sampson et al., 2014).

Three CRISPR-Cas components have been shown to repress gene expression to facilitate virulence in *F. novicida*: Cas9, tracrRNA and the newly-described small-CRISPR- associated scaRNA (Sampson et al., 2013). These elements are shown to regulate expression of the FTN_1103 gene, which codes for a bacterial lipoprotein (BLP) (Sampson et al., 2013). In the absence of any of these factors (but not other Cas proteins or the crRNA), expression of the FTN_1103 mRNA coding for a BLP is upregulated (Sampson et al., 2013). The increase in BLP synthesis was shown to disturb bacterial envelope integrity, which results in TLR2 and AIM2/ASC activation. This activation is likely a consequence of DNA escaping from the bacteria, which would activate both these receptors. As a result, BLP overproduction promotes inflammation and attenuates *F. novicida* virulence (Sampson et al., 2014). Interestingly, the double mutant lacking *cas9* and FTN_1103 only partially restores virulence, suggesting there are additional factors that are regulated by Cas9 (Sampson et al., 2014).

The mechanism by which expression of FTN_1103 is upregulated by Cas9 remains unknown. Cas9, tracrRNA and scaRNA are part of the same mechanism as a triple deletion mutant shows the same effect as any of the single mutants (Sampson et al., 2013). scaRNA is encoded next to the CRISPR array in what appears to be a degenerated repeat. It is transcribed from a promoter in a former spacer and terminates within the next spacer, covering a complete degenerated repeat (Figure 11) (Chylinski et al., 2014). Furthermore, scaRNA is predicted to base-pair with tracrRNA, reminiscent of tracrRNA:crRNA base-pairing (Chylinski et al., 2014). These predictions are supported by point mutation analysis where substitutions in regions predicted to mediate the interaction have a similar effect than deleting any of the three components. In addition, repression of FTN_1103 was partially restored by compensatory substitutions that would regenerate the tracrRNA:scaRNA interaction (Sampson et al., 2013). tracrRNA was also predicted to base-pair with the FTN_1103 mRNA in a region close to its RBS.

Therefore, it was proposed that the tracrRNA:scaRNA duplex binds Cas9 and guides it to its target (FTN_1103 mRNA) promoting its degradation. However, the details of how Cas9 is guided and how the target is degraded or destabilized are unclear. In *F. novicida* Cas9 (FnoCas9), the HNH and RuvC-like domains, as well as the R-rich motif, are conserved. Substitutions in conserved amino acids of the nuclease domains (D11A and H969A for HNH and RuvC-like, respectively), which are essential for Cas9 activity in *S. pyogenes* (Jinek et al., 2012), do not affect FnoCas9-mediated

regulation of FTN_1103, indicating that the nuclease activity of Cas9 is not important for FTN_1103 regulation. However, the R59A substitution, located in the R-rich motif, shows the same effect than *cas9* deletion. Furthermore, scaRNA, tracrRNA and the FTN_1103 mRNA were shown to co-immunoprecipitate with WT Cas9 but not with the R59A mutant, and FTN_1103 mRNA was shown to be less stable in the WT than in the Cas9 mutant (Sampson et al., 2013).

The proposed model includes 3 steps: 1) tracrRNA:scaRNA:Cas9 complex formation, 2) targeting of FTN1103 mRNA by the complex through base-pair complementarity between tracrRNA and FTN1103 mRNA, 3) destabilization and degradation of the target mRNA by an unknown mechanism.

Results

The following work was done in collaboration with the group of David Weiss at Emory University in Atlanta, USA. The complete results are included in the manuscript entitled *"Catalytically active Cas9 mediates transcriptional interference to facilitate bacterial virulence",* which at the moment of writing this thesis manuscript is under revision in the journal *Molecular Cell*. In this section, some additional results are described together with the summary of the main findings and conclusions. Please refer to the full manuscript currently in revision in *Molecular Cell* in the appendix section of this thesis manuscript for the figures and complete results.

F. novicida (Fno)Cas9 binds and cleaves DNA specifically *in vitro*

As mentioned above, tracrRNA, scaRNA and FnoCas9, negatively affect mRNA FTN_1103 expression by an unknown mechanism (Sampson et al., 2013). The aim of this study is to investigate the mechanism by which FnoCas9 regulates gene expression. With this in mind, we used purified FnoCas9 to analyze its biochemical properties *in vitro*. Though the functionality of FnoCas9 in immunity remains to be determined, its nuclease activity on the target DNA has previously been shown *in vitro* (Fonfara, I. et al. 2013). As a control to demonstrate that the purified FnoCas9 was catalytically active in our experimental conditions, we analyzed the binding and cleavage activities on a DNA fragment containing the target sequence. As expected, the Cas9:tracRNA:crRNA complex bound and cleaved a DNA target specifically, as shown by electrophoretic mobility shift assay (EMSA) and plasmid cleavage assay,

respectively (Figure 9. A-B). These results confirmed that our purified FnoCas9 was active in the tested conditions.

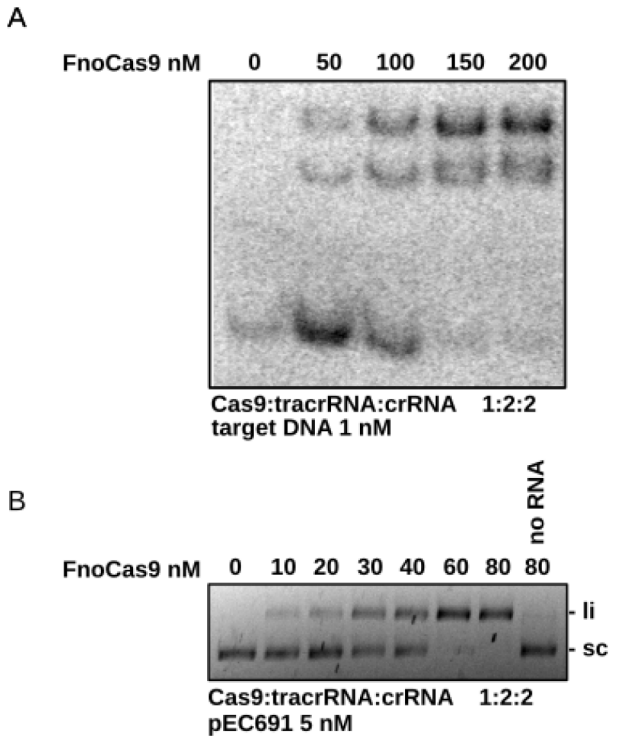

Figure 9. FnoCas9 binds to and cleaves its target DNA. A) EMSA showing FnoCas9 binding to a DNA fragment containing the target sequence in the presence of tracrRNA and a crRNA B) Cleavage assay showing the specific DNA nuclease activity of FnoCas9. The supercoiled (sc) and linear (li) forms of the plasmid are indicated. The linear form is only observed in the presence of the tracrRNA:crRNA duplex and FnoCas9. The last lane contains no RNA duplex.

FnoCas9 binds tracrRNA:crRNA and tracrRNA:scaRNA *in vitro*

The fact that the FnoCas9 DNA nuclease activity is conserved suggests that the type II-B CRISPR-Cas system is also active in defense against DNA invasion, implying that FnoCas9 is able to bind both tracrRNA:crRNA and tracrRNA:scaRNA duplexes. In order to test this, we performed EMSA by incubating FnoCas9 with either tracrRNA:crRNA or tracrRNA:scaRNA preformed duplexes. As expected, Cas9 was able to bind both duplexes (manuscript Fig. S5), suggesting it might have dual function in defense and regulation.

FnoCas9 specific binding to its potential RNA targets is not detected

In order to investigate the mechanism by which FnoCas9 regulates FTN_1103, we evaluated its affinity to FTN_1103 mRNA by EMSA. To this end, we produced a fragment of the FTN_1103 transcript that contained the proposed interaction site with tracrRNA. However, while FnoCas9 binds to FTN_1103 in the absence of the duplex

RNA (Figure 10 A), the affinity did not increase upon addition of either tracrRNA, scaRNA or the performed duplex (Figure 10 B-D) suggesting that the binding to this RNA fragment is unspecific.

To test the possibility that FnoCas9 binds in a different location of the transcript, we performed EMSA on the full-length FTN_1103 transcript, which includes FTN_1104 (manuscript Figs. 1A and 1B). However, we failed to detect any specific binding of FnoCas9 to the mRNA (data not shown).

RNAseq differential expression analysis comparing gene expression of the WT strain to deletion mutants of scaRNA, tracrRNA and Cas9 detected that FTN_1101 was regulated in addition to the FTN_1103-FTN_1104 transcript (manuscript Figs. 1A and 1B). These results were also confirmed by Northern blot analyses (manuscript Fig. 1C-E). However, FnoCas9 did not bind specifically to the FTN_1101 transcript either (data not shown). Together, these results suggested that FnoCas9-mediated gene regulation might require additional factors or different conditions.

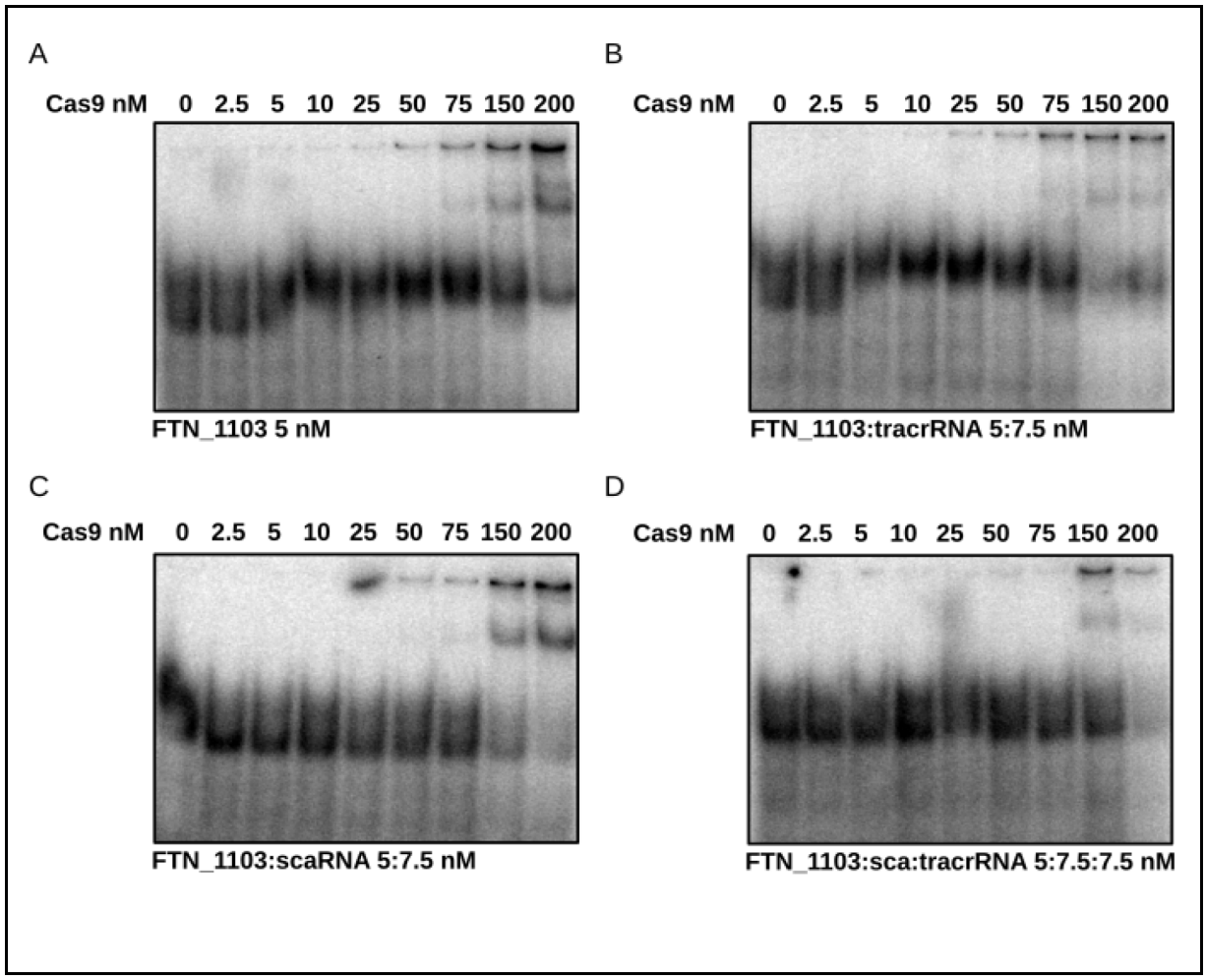

Figure 10 FnoCas9 specific binding to its potential RNA targets is not detected. EMSA showing affinity of *F. novicida* Cas9 for isotopically labelled FTN_1103 mRNA by itself (A) or in the presence of scaRNA (B), tracrRNA (C) or scaRNA and tracrRNA (D). tracrRNA and scaRNA were used (Fig 1. B and C) in a concentration 1.5-fold higher than FTN_1103 mRNA. RNAs were incubated at 95 ℃ in Binding Buffer and let cool down slowly to room temperature. Then increasing concentrations of Cas9 were added and the binding was analyzed by electrophoresis in polyacrylamide 6 % with 20 mM Tris-acetate pH 8.5.

FnoCas9 regulates its target genes via a conserved sequence in the 5′ UTR that is complementary to scaRNA

The absence of specific binding of FnoCas9 to its proposed target mRNAs *in vitro* led us to re-examine the regulating mechanism. In order to investigate the regions that are essential for gene regulation by FnoCas9, the 5′ UTR sequences of FTN_1101 and FTN_1104 were aligned. Interestingly, a region of 17 bp that was identical between the two sequences was identified (manuscript Fig. 2A). Furthermore, a reporter fusion was regulated by FnoCas9 when the 5′ UTR of any of the two genes was introduced downstream of the promoter, regardless of the promoter that was used (manuscript Fig. 2C-D).

In addition, 11 consecutive nt of scaRNA were predicted to base pair with the conserved sequences of FTN_1104 and FTN_1101 (manuscript Figs. 3A and B), suggesting that targeting might be mediated by scaRNA and not by tracrRNA.

FnoCas9 interacts with the DNA of the regulated genes in a PAM-dependent manner

Further inspection of the adjacent sequences localized a putative PAM (TGG) sequence downstream of these regions (Figs. 3A and B). Moreover, substitutions that disrupted the putative PAM abrogated regulation of the reporter fusions by FnoCas9 (manuscript Figs. 3C and 3D). Since the PAM is encoded in the non-target strand, these results indicated that FnoCas9 regulates its targets by binding to the DNA. Indeed, EMSA experiments showed that FnoCas9 interacted with a DNA target containing 11 bp of complementarity to scaRNA (manuscript Fig. 3E). Furthermore, this interaction was only observed in the presence of the PAM and the

tracrRNA:scaRNA duplex (manuscript Fig. 3E).

The number of base pairs between scaRNA and the target DNA determines the level of transcriptional repression

To evaluate the extent of scaRNA-DNA complementarity that FnoCas9 requires for gene regulation, we tested the expression of reporter fusions containing sequences with different numbers of base pairs to scaRNA (manuscript Fig. S3A). Higher complementarity of the target to scaRNA correlated with higher levels of repression of the reporter fusions (manuscript Fig 4A-F). In addition, the number of basepairs with scaRNA was also reflected in the affinity of FnoCas9 to the target DNA *in vitro,* the higher the complementarity the higher the affinity (manuscript Fig. 4A-F).

scaRNA can mediate cleavage of complementary target DNA

Interestingly, transformation efficiency of a plasmid containing 20 nt of complementarity to scaRNA was drastically reduced compared to plasmid with less complementarity, indicating that scaRNA can mimic crRNA function provided that there is enough complementarity with the target DNA (manuscript Fig. 4F).

Transcription interference by FnoCas9 requires binding to a region in close proximity to the promoter

Next, we investigated whether the distance of the target sequence to the promoter affected FnoCas9-mediated regulation. To this end, we tested the expression of reporter fusions with increasing distance to the promoter (manuscript Fig. 4H). While binding of FnoCas9 to the target was unaltered regardless of the distance to the promoter, repression was abrogated when the target sequence was placed 20 bp downstream of the transcription start site (manuscript Fig. 4H-K). This suggested that FnoCas9 binding was able to inhibit transcription initiation but not elongation.

FnoCas9 forms two distinct complexes in the cell containing scaRNA or crRNA

We hypothesized that two subpopulations of FnoCas9-complexes coexist in the cell, each containing one of the RNA duplexes. To test whether the levels of one of the duplexes influenced the abundance of the other, we performed Northern blot analyses in the WT and the deletion mutants for tracrRNA, scaRNA and crRNA and measured the abundance of the three ncRNAs. While the absence of scaRNA had no observable

effect on crRNA abundance, deleting crRNA caused a significant increase in scaRNA abundance (manuscript Fig. 5A-C). Furthermore, this increase in scaRNA abundance translated into a higher repression of the target genes (manuscript Fig. 5D-H). This indicated that FnoCas9 binding to scaRNA had a stabilizing effect. Moreover, these results suggest that crRNA and scaRNA seem to compete with crRNA for binding Cas9.

Engineering scaRNA allows artificial regulation of desired genes

Finally, to prove that this system can be used to repress other genes, we modified scaRNA to target the polymyxin resistance genes (FTN_0544 and FTN_0545) (manuscript Fig. 7A). As expected, targeting a region near the promoter of these genes resulted in a lower gene expression and a reduction in the resistance to this antibiotic (manuscript Fig. 7B-E).

Discussion

In this study, we showed that scaRNA guides FnoCas9 to bind downstream of the promoter of its target genes. Binding of FnoCas9 to the DNA interferes with transcription, repressing gene expression. Though FnoCas9 is able to cleave DNA, the limited complementarity of scaRNA with the DNA prevents digestion of the chromosome.

Taken together, the results shown in this study suggest that FnoCas9 has a dual functionality. On the one hand, when bound to the tracrRNA:crRNA duplex, it maintains the classical immunity function. On the other, it has evolved the capability to regulate gene expression together with tracrRNA:scaRNA (Figure 11).

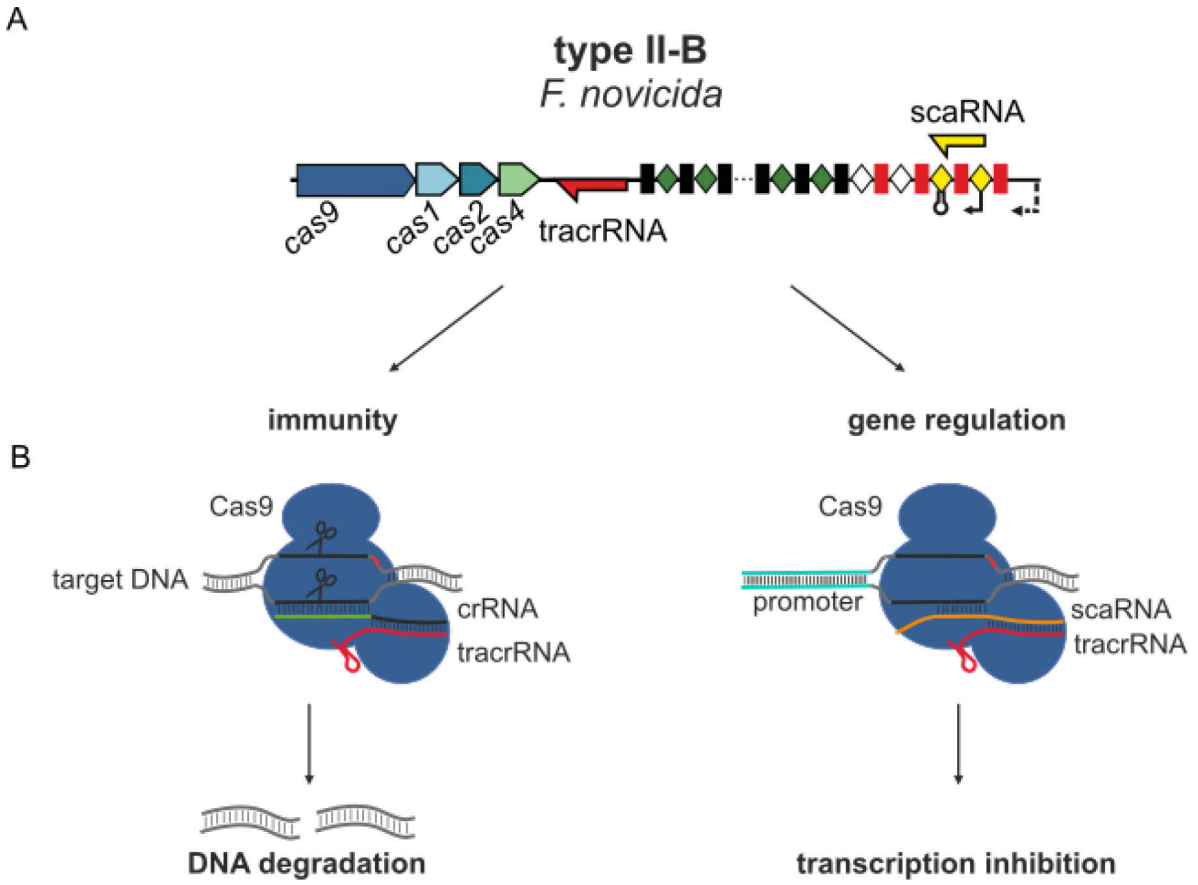

Figure 11 Scheme of the immunity and gene regulation mechanisms by the type II-B CRISPR-Cas system of *F. novicida*. A) Representation of the CRISPR-Cas locus. Yellow and red arrows represent scaRNA and tracrRNA, respectively. Black squares and green diamonds indicate repeats and spacers, respectively. Putative degenerated repeats are shown as red squares, intercalated by spacers (white diamonds). Spacers that evolved to constitute scaRNA are shown in yellow. The confirmed and putative promoters of scaRNA are represented with solid or dotted bent arrows, respectively. The *cas* genes are indicated. B) Mechanisms of immunity and regulation. FnoCas9 (blue) is guided by the crRNA:tracrRNA duplex to target and cleave DNA, constituting the immunity pathway (left). scaRNA:tracrRNA guides FnoCas9 to the regions adjacent to the promoters of the target genes, repressing transcription (right). Limited complementarity with scaRNA prevents cleavage. This figure is adapted and modified from (Chylinski et al., 2014).

The mechanism discovered here, opens the possibility that regulatory functions of Cas9 are more widespread than previously thought. The fact that Cas9 is able to regulate gene expression while maintaining its role in immunity, means that evolution of this new functionality would not necessarily entail any evolutionary costs. It has been proposed that scaRNA evolved from degeneration the CRISPR array (Chylinski et al.,

2014). A possible pathway for the evolution of this regulatory functionality might start with the acquisition of self-targeting spacers. Mutations in the spacers that prevent them from cleaving the DNA would allow the cell to survive. If the target region of this spacer is in close proximity to a promoter, it would repress the downstream gene. Evolution of an independent promoter driving transcription of this spacer would then allow regulation of the newly evolved function, independently of the expression of the immunity components. A bioinformatic search of partially self-targeting spacers with an independent promoter (and a terminator) should identify CRISPR-Cas systems with putative regulatory functions.

In the case of *F. novicida,* repression of the BLPs is clearly advantageous, as it allows the cells to evade the immune system (Sampson et al., 2014). Yet, it is likely that under different conditions, BLP production is necessary. Future studies should investigate how the regulatory activity of FnoCas9 is regulated. It is likely that it is achieved by regulating scaRNA expression, since this would allow maintaining the immunity function of FnoCas9 unaltered.

The newly discovered mechanism might also have implications regarding off target effects of Cas9. Traditionally, off-targets are defined as DSB in undesired locations. Our results show that limited complementarity between a guide RNA and the DNA is sufficient to repress transcription. Though the level of repression is likely to depend on the expression levels of Cas9 and the target DNA, this form of off-target effect might be significant in some conditions and should be considered.

Materials and Methods

For a detailed description of the experimental procedures, please refer to the manuscript in the appendix. In this section, only additional experiments (not shown in the manuscript) are described.

Table 6. List of primers used in the study.

Code	Sequence 5′-3′	F/R[a]	Usage[b]	Reference
Northern blot assays				
OLEC4268	CCAACAGCCTGCCTGCCACT	R	NB FTN_1101	This study
OLEC4269	GTCTCCGTACTGTAACTTGC	R	NB FTN_RS05660	This study
OLEC4274	ACCCAACTCACCATCGCCACA	R	NB FTN_1103	This study
OLEC4276	TGGCAGCCAAAATGCGGTAG	R	NB FTN_1104	This study
OLEC4323	ACCTACAGACCCTTTACGCC	R	NB 16S rRNA	This study

OLEC2861	CTAACAGTAGTTTACCAAATAATTCAGCAACTGAAAC	R	NB crRNA	This study
OLEC2866	ATTACAGAGCATTAATTATTTGGTACATTTATAATTT	R	NB tracrRNA	This study
OLEC2864	AACACAAGTACATACCAAATAATCTAACAACTGAAAC	R	NB scRNA	This study
OLEC5225	ACCTACTTTCACCTGGGCAA	R	NB 5S rRNA	This study
RNA production				
OLEC4211	TAATACGACTCACTATA	F	IVT T7 promoter	This study
OLEC4407	AAAAATAAGTAGGTCTAAAAGTGAATTTTCTAGCTACTTTAAACTAACACAAGTACATACCAAATAATCTAACAACCTATAGTGAGTCGTATTA	R	IVT T7 scaRNA template with OLEC4211	This study
OLEC3089	GAAATTAATACGACTCACTATAGGATAACTCAATTTGTAAAAAAGTTTCAGTTGCTGAATTATTTGGTAAAC	F	IVT T7 crRNA speM spacer	(Fonfara et al., 2013)
OLEC3090	GTTTACCAAATAATTCAGCAACTGAAACTTTTTTACAAATTGAGTTATCCTATAGTGAGTCGTATTAATTTC	R	IVT T7 crRNA speM spacer	(Fonfara et al., 2013)
OLEC3102	GAAATTAATACGACTCACTATA GGGTACCAAATAATTAATGCTCTG	F	IVT T7-tracrRNA (processed)	(Fonfara et al., 2013)
OLEC3103	GTTATTCAGACGTGTCAAACAG	R	IVT T7-tracrRNA	(Fonfara et al., 2013)
OLEC4649	TAATACGACTCACTATAGGATCTAAAATTATAAATGTACCAAATAATTAATGC	F	IVT T7-tracrRNA (unprocessed)	This study
OLEC4683	AAAAATAAGTAGGTCTAAAAGTGAATTTTCTAGCTACTTTAAACTAACACAAGTACATACCAAATAATCTAACAACTGAAACTTACAGCTAATCTTTTTCATTTGCTCCTATAGTGAGTCGTATTA	R	IVT T7-scaRNA (long) with OLEC4211	
DNA EMSAs				
OLEC9025	TAAATACCAATTGACATATATAAATGATTCTGATATAAATTAGATAAGGGAGGCCAACACTTGTCACTACTCTGACG	F	DNA EMSA 0 bp complementarity)	This study
OLEC9026	CGTCAGAGTAGTGACAAGTGTTGGCCTCCCTTATCTAATTTATATCAGAATCATTTATATATGTCAATTGGTATTTA	R	DNA EMSA 0 bp complementarity	This study
OLEC9027	TAAATACCAATTGACATATATAAATGATTCTGATATAAATTAGATAAGGGAGCTGTAATGGGGCCAACACTTGTCACTACTCTGACG	F	DNA EMSA 8 bp complementarity	This study
OLEC9028	CGTCAGAGTAGTGACAAGTGTTGGCCCCATTACAGCTCCCTTATCTAATTTATATCAGAATCATTTATATATGTCAATTGGTATTTA	R	DNA EMSA 8 bp complementarity	This study
OLEC9029	TAAATACCAATTGACATATATAAATGATTCTGATATAAATTAGATAAGGGATTAGCTGTAATGGGGCCAACACTTGTCACTACTCTGACG	F	DNA EMSA 11 bp complementarity	This study
OLEC9030	CGTCAGAGTAGTGACAAGTGTTGGCCCCATTACAGCTAATCCCTTATCTAATTTATATCAGAATCATTTATATATGTCAATTGGTATTTA	R	DNA EMSA 11 bp complementarity	This study
OLEC9031	TAAATACCAATTGACATATATAAATGATTCTGATATAAATTAGATAAGGGAAAGATTAGCTGTAATGGGGCCAACACTTGTCACTACTCTGACG	F	DNA EMSA 15 bp complementarity	This study
OLEC9032	CGTCAGAGTAGTGACAAGTGTTGGCCCCATTACAGCTAATCTTTCCCTTATCTAATTTATATCAGAATCATTTATATATGTCAATTGGTATTTA	R	DNA EMSA 15 bp complementarity	This study
OLEC9033	TAAATACCAATTGACATATATAAATGATTCTGATATAAATTAGATAAGGGATGAAAAAGATTAGCTGTAATGGGGCCAACACTTGTCACTACTCTGACG	F	DNA EMSA 20 bp complementarity	This study
OLEC9034	CGTCAGAGTAGTGACAAGTGTTGGCCCCATTACAGCTAATCTTTTTCATCCCTTATCTAATTTATATCAGAATCATTTATATATGTCAATTGGTATTTA	R	DNA EMSA 20 bp complementarity	This study

OLEC9035	TAAATACCAATTGACATATATAAATGATTCTGAT ATAAATTAGATAAGGGTTAGTATTAGCTGTAATG GGGCCAACACTTGTCACTACTCTGACG	F	DNA EMSA 11 bp complementarity 5 bp from TSS	This study
OLEC9036	CGTCAGAGTAGTGACAAGTGTTGGCCCCATTACA GCTAATACTAACCCTTATCTAATTTATATCAGAA TCATTTATATATGTCAATTGGTATTTA	R	DNA EMSA 11 bp complementarity 5 bp from TSS	This study
OLEC9037	TAAATACCAATTGACATATATAAATGATTCTGAT ATAAATTAGATAAGGGTTTACTTAGTATTAGCTG TAATGGGGCCAACACTTGTCACTACTCTGACG	F	DNA EMSA 11 bp complementarity 10 bp from TSS	This study
OLEC9038	CGTCAGAGTAGTGACAAGTGTTGGCCCCATTACA GCTAATACTAAGTAAACCCTTATCTAATTTATAT CAGAATCATTTATATATGTCAATTGGTATTTA	R	DNA EMSA 11 bp complementarity 10 bp from TSS	This study
OLEC9039	TAAATACCAATTGACATATATAAATGATTCTGAT ATAAATTAGATAAGGGAGAAATCCAATTTACTTA GTATTAGCTGTAATGGGGCCAACACTTGTCACTA CTCTGACG	F	DNA EMSA 11 bp complementarity 20 bp from TSS	This study
OLEC9040	CGTCAGAGTAGTGACAAGTGTTGGCCCCATTACA GCTAATACTAAGTAAATTGGATTTCTCCCTTATC TAATTTATATCAGAATCATTTATATATGTCAATT GGTATTTA	R	DNA EMSA 11 bp complementarity 20 bp from TSS	This study

Plasmid cleavage assay

FnoCas9 was incubated with pEC691 (Fonfara et al., 2014) (containing the protospacrer and the PAM) for 1 h at 37 °C in the presence of the pre-formed tracrRNA:crRNA duplex. Reaction was done in KGB buffer (100 mM K-glutamate, 25 mM Tris-acetate pH 7.5, 10 mM Mg-acetate, 10 ug/ml BSA, 0.5 mM ß-Mercaptoethanol). After incubation, reaction was stopped by adding 3 µl stopping solution and samples were resolved by electrophoresis in a 0.8% agarose gel.

DNA EMSA

A tracrRNA:crRNA pre-formed duplex was incubated with FnoCas9 for 15 min at 37 °C. Next, 1 nM target DNA was added (1:5 cold:hot) and incubated for 1 h at 37 °C. Incubation was performed in Ca^{+2}- containing binding buffer (20 mM tris-HCl pH 7.5, 100 mM KCl, 5 mM $CaCl^{+2}$, 5% Glycerol, 1 mM DTT). Samples were analyzed by PAGE in 6% PAA running buffer (0.5 x TBE 8pH, 5 mM $CaCl^{+2}$).

RNA EMSA

FTN_1103 RNA transcript was produced by *in vitro* transcription using the AmpliScribe™ T7-Flash™ Transcription Kit (Epicentre) according to the manufacturer's instructions. The DNA template was produced by PCR using oligos (OLEC4405; TAATACGACTCACTATAGGATGCCCTGACTGCTCTCGTG and OLEC4406; CTTGCCACAACTGCCCAATATCAC). 1 nM of isotopically labelled

FTN_1103 RNA was incubated in 10 µl reactions with increasing concentrations of Cas9 either by itself or in the presence of scaRNA, tracrRNA, or tracrRNA:scaRNA in concentration 1.5-fold higher than FTN_1103 mRNA. For pre-annealing, tracrRNA:scaRNA were incubated at 95 °C in RNA annealing buffer (1 M NaCl, 100 mM HEPES, pH 7.5) and let cool down slowly to room temperature. Then increasing concentrations of Cas9 were added and the binding was analyzed by electrophoresis in polyacrylamide 6 % with 20 mM Tris-acetate pH 8.5.

Contributions

The work in this chapter was done in collaboration with the group of David Weiss at Emory University in Atlanta, USA. The experiments in *F. novicida,* including analysis of gene expression by qRT-PCR, were performed by the Weiss group. RNAseq analysis and the Northern Blot of the sRNAs were done by Anaïs Le Rhun in our laboratory. The study of Cas9 activity *in vitro* and the Northern Blot analysis to evaluate gene expression of Cas9 targets were done by myself. The experimental design and data analysis was performed jointly by the Weiss and the Charpentier groups (see also the “Authors contributions” section in the manuscript).

References

Anders, C., Niewoehner, O., Duerst, A., and Jinek, M. (2014). Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature *advance online publication*.

Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nat. Biotechnol. *34*, 933–941.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science *315*, 1709–1712.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiol. Read. Engl. *151*, 2551–2561.

Briner, A.E., Donohoue, P.D., Gomaa, A.A., Selle, K., Slorach, E.M., Nye, C.H., Haurwitz, R.E., Beisel, C.L., May, A.P., and Barrangou, R. (2014). Guide RNA functional modules direct Cas9 activity and orthogonality. Mol. Cell *56*, 333–339.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and Oost, J. van der (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. Science *321*, 960–964.

Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. *10*, 726–737.

Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E.V. (2014). Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res. *42*, 6091–6105.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. Science *339*, 819–823.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

Dominguez, A.A., Lim, W.A., and Qi, L.S. (2016). Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nat. Rev. Mol. Cell Biol. *17*, 5–15.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science *346*, 1258096.

Dugar, G., Herbig, A., Förstner, K.U., Heidrich, N., Reinhardt, R., Nieselt, K., and Sharma, C.M. (2013). High-Resolution Transcriptome Maps Reveal Strain-Specific Regulatory Features of Multiple Campylobacter jejuni Isolates. PLoS Genet *9*, e1003495.

Dugar, G., Leenay, R.T., Eisenbart, S.K., Bischler, T., Aul, B.U., Beisel, C.L., and Sharma, C.M. (2018). CRISPR RNA-Dependent Binding and Cleavage of Endogenous RNAs by the Campylobacter jejuni

Cas9. Mol. Cell *69*, 893-905.e7.

Fonfara, I., Rhun, A.L., Chylinski, K., Makarova, K.S., Lécrivain, A.-L., Bzdrenga, J., Koonin, E.V., and Charpentier, E. (2013). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. gkt1074.

Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K.S., Lécrivain, A.-L., Bzdrenga, J., Koonin, E.V., and Charpentier, E. (2014). Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nucleic Acids Res. *42*, 2577–2590.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. *109*, E2579–E2586.

Heler, R., Samai, P., Modell, J.W., Weiner, C., Goldberg, G.W., Bikard, D., and Marraffini, L.A. (2015). Cas9 specifies functional viral targets during CRISPR-Cas adaptation. Nature *advance online publication*.

Hille, F., Richter, H., Wong, S.P., Bratovič, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. Cell *172*, 1239–1259.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M., and Nakata, A. (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. J. Bacteriol. *169*, 5429–5433.

Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C., and Brouns, S.J.J. (2017). CRISPR-Cas: Adapting to change. Science *356*, eaal5056.

Jansen, R., Embden, J.D.A. van, Gaastra, W., and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. Mol. Microbiol. *43*, 1565–1575.

Jiang, W., Bikard, D., Cox, D., Zhang, F., and Marraffini, L.A. (2013). RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat. Biotechnol. *31*, 233–239.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816–821.

Jinek, M., East, A., Cheng, A., Lin, S., Ma, E., and Doudna, J. (2013). RNA-programmed genome editing in human cells. ELife *2*, e00471.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., et al. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. Science *343*, 1247997.

Ka, D., Jang, D.M., Han, B.W., and Bae, E. (2018). Molecular organization of the type II-A CRISPR

adaptation module and its interaction with Cas9 via Csn2. Nucleic Acids Res. *46*, 9805–9815.

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9-histone demethylase fusion. Nat. Methods *12*, 401–403.

Kieper, S.N., Almendros, C., Behler, J., McKenzie, R.E., Nobrega, F.L., Haagsma, A.C., Vink, J.N.A., Hess, W.R., and Brouns, S.J.J. (2018). Cas4 Facilitates PAM-Compatible Spacer Selection during CRISPR Adaptation. Cell Rep. *22*, 3377–3384.

Kingry, L.C., and Petersen, J.M. (2014). Comparative review of Francisella tularensis and Francisella novicida. Front. Cell. Infect. Microbiol. *4*.

Koonin, E.V., Makarova, K.S., and Zhang, F. (2017). Diversity, classification and evolution of CRISPR-Cas systems. Curr. Opin. Microbiol. *37*, 67–78.

Lee, H., Zhou, Y., Taylor, D.W., and Sashital, D.G. (2018). Cas4-Dependent Prespacer Processing Ensures High-Fidelity Programming of CRISPR Arrays. Mol. Cell *70*, 48-59.e5.

Louwen, R., Horst-Kreft, D., Boer, A.G. de, Graaf, L. van der, Knegt, G. de, Hamersma, M., Heikema, A.P., Timms, A.R., Jacobs, B.C., Wagenaar, J.A., et al. (2013). A novel link between Campylobacter jejuni bacteriophage defense, virulence and Guillain–Barré syndrome. Eur. J. Clin. Microbiol. Infect. Dis. *32*, 207–226.

Louwen, R., Staals, R.H.J., Endtz, H.P., Baarlen, P. van, and Oost, J. van der (2014). The Role of CRISPR-Cas Systems in Virulence of Pathogenic Bacteria. Microbiol. Mol. Biol. Rev. *78*, 74–88.

Ma, K., Cao, Q., Luo, S., Wang, Z., Liu, G., Lu, C., and Liu, Y. (2018). cas9 Enhances Bacterial Virulence by Repressing the regR Transcriptional Regulator in Streptococcus agalactiae. Infect. Immun. *86*, e00552-17.

Makarova, K.S., Grishin, N.V., Shabalina, S.A., Wolf, Y.I., and Koonin, E.V. (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. Biol. Direct *1*, 7.

Makarova, K.S., Haft, D.H., Barrangou, R., Brouns, S.J.J., Charpentier, E., Horvath, P., Moineau, S., Mojica, F.J.M., Wolf, Y.I., Yakunin, A.F., et al. (2011). Evolution and classification of the CRISPR–Cas systems. Nat. Rev. Microbiol. *9*, 467–477.

Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2013). Comparative genomics of defense systems in archaea and bacteria. Nucleic Acids Res. *41*, 4360–4377.

Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., et al. (2015). An updated evolutionary classification of CRISPR-Cas systems. Nat. Rev. Microbiol. *13*, 722–736.

Mali, P., Esvelt, K.M., and Church, G.M. (2013). Cas9 as a versatile tool for engineering biology. Nat. Methods *10*, 957–963.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. Science *322*, 1843–1845.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. J. Mol. Evol. *60*, 174–182.

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. Cell.

Nuñez, J.K., Kranzusch, P.J., Noeske, J., Wright, A.V., Davies, C.W., and Doudna, J.A. (2014). Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nat. Struct. Mol. Biol. *advance online publication*.

Nuñez, J.K., Lee, A.S.Y., Engelman, A., and Doudna, J.A. (2015a). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *advance online publication*.

Nuñez, J.K., Lee, A.S.Y., Engelman, A., and Doudna, J.A. (2015b). Integrase-mediated spacer acquisition during CRISPR–Cas adaptive immunity. Nature *519*, 193–198.

Ratner, H.K., Sampson, T.R., and Weiss, D.S. (2015). I can see CRISPR now, even when phage are gone: a view on alternative CRISPR-Cas functions from the prokaryotic envelope. Curr. Opin. Infect. Dis. *28*, 267–274.

Rousseau, B.A., Hou, Z., Gramelspacher, M.J., and Zhang, Y. (2018). Programmable RNA Cleavage and Recognition by a Natural CRISPR-Cas9 System from Neisseria meningitidis. Mol. Cell *69*, 906-914.e4.

Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.-L., and Weiss, D.S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature *497*, 254–257.

Sampson, T.R., Napier, B.A., Schroeder, M.R., Louwen, R., Zhao, J., Chin, C.-Y., Ratner, H.K., Llewellyn, A.C., Jones, C.L., Laroui, H., et al. (2014). A CRISPR-Cas system enhances envelope integrity mediating antibiotic resistance and inflammasome evasion. Proc. Natl. Acad. Sci. 201323025.

Shiimori, M., Garrett, S.C., Graveley, B.R., and Terns, M.P. (2018). Cas4 Nucleases Define the PAM, Length, and Orientation of DNA Fragments Integrated at CRISPR Loci. Mol. Cell *70*, 814-824.e6.

Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. Mol. Cell *60*, 385–397.

Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I., et al. (2017). Diversity and evolution of class 2 CRISPR–Cas systems. Nat. Rev. Microbiol. *15*, 169–182.

Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A., and Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. Nat. Commun. *7*, 12778.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by

the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62–67.

Szczelkun, M.D., Tikhomirova, M.S., Sinkunas, T., Gasiunas, G., Karvelis, T., Pschera, P., Siksnys, V., and Seidel, R. (2014). Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc. Natl. Acad. Sci. 201402597.

Wei, Y., Chesne, M.T., Terns, R.M., and Terns, M.P. (2015). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. Nucleic Acids Res. Gku1407.

Westra, E.R., Buckling, A., and Fineran, P.C. (2014). CRISPR-Cas systems: beyond adaptive immunity. Nat. Rev. Microbiol. *advance online publication*.

Zhang, Q., Doak, T.G., and Ye, Y. (2013). Expanding the catalog of cas genes with metagenomes. Nucleic Acids Res. gkt1262.

Appendix

Manuscript

Catalytically Active Cas9 Mediates Transcriptional Interference to Facilitate Bacterial Virulence

Hannah K. Ratner[1,2,3], Andrés Escalera-Maurer[4,5], Anaïs Le Rhun[4,5], Siddharth Jaggavarapu[2,3,6], Jessie E. Wozniak[1,2,3], Emily K. Crispell[1,2,3], Emmanuelle Charpentier[4,5,7], David S. Weiss [1,2,3,6]

[1]Microbiology and Molecular Genetics Program, Emory University, Atlanta, GA 30329.

[2]Emory Vaccine Center, Emory University, Atlanta, GA 30329.

[3]Yerkes National Primate Research Center, Emory University, Atlanta, GA 30329.

[4]Max Planck Unit for the Science of Pathogens, D-10117 Berlin, Germany

[5]Helmholtz Centre for Infection Research, Department of Regulation in Infection Biology, D-38124, Braunschweig, Germany

[6]Division of Infectious Diseases, Department of Medicine, Emory University School of Medicine, Atlanta, Georgia 30329.

[7]Institute for Biology, Humboldt University, D-10115 Berlin,

Germany *Corresponding author/Lead contact: David S. Weiss

Emory Vaccine Center

954 Gatewood Rd, Room 2028

Atlanta, GA 30329

Tel: (404) 727-8214

Email: david.weiss@emory.edu

Please refer to the online version of the manuscript:

Ratner, Hannah K. and Escalera-Maurer, Andrés and Le Rhun, Anaïs and Jaggavarapu, Siddarth and Wozniak, Jessie E. and Crispell, Emily K. and Charpentier, Emmanuelle and Weiss, David S., Catalytically Active Cas9 Mediates Transcriptional Interference to Facilitate Bacterial Virulence (April 16, 2019).

Available at SSRN:
https://ssrn.com/abstract=3372971orhttp://dx.doi.org/10.2139/ssrn.3372971